Organic Matter Fluorescence in Municipal Water Recycling Schemes: Toward a Unified PARAFAC Model

Kathleen R. Murphy,* Adam Hambly, Sachin Singh, Rita K. Henderson, Andy Baker, Richard Stuetz, and Stuart J. Khan

School of Civil and Environmental Engineering, The University of New South Wales, Sydney NSW 2052, Australia

S *Supporting Information*

ABSTRACT: Organic matter (OM) is a ubiquitous constituent of natural waters quantifiable at very low levels using fluorescence spectroscopy. This technique has recognized potential in a range of applications where the ability to monitor water quality in real time is desirable, such as in water treatment systems. This study used PARAFAC to characterize a large ($n = 1479$) and diverse excitation emission matrix (EEM) data set from six recycled water treatment plants in Australia, for which sources of variability included geography, season, treatment processes, pH and fluorometer settings. Five components were identified independently in four or more plants, none of which were generated during the treatment process nor were typically entirely removed. PARAFAC scores could be obtained from EEMs by simple regression. The results have important implications for online monitoring of OM fluorescence in treatment plants, affecting choices regarding experimental design, instrumentation and the optimal wavelengths for tracking fluorescent organic matter through the treatment process. While the multimodel comparisons provide a compelling demonstration of PARAFAC's ability to distill chemical information from EEMs, deficiencies identified through this process have broad implications for interpreting and reusing (D)OM-PARAFAC models.

INTRODUCTION

Research findings over the past decades have highlighted the benefits of monitoring organic matter (OM) fluorescence to evaluate water quality in natural[1,2] and engineered[3,4] aquatic systems. Produced by the breakdown of terrestrial and aquatic plant matter to humic, fulvic, and amino acids, organic matter is found in all natural waters. Dissolved (DOM) and colloidal forms are often quantified indirectly using excitation emission matrix fluorescence spectroscopy, in which samples are excited at a range of wavelengths and emission recorded across a range of wavelengths, generating three-dimensional intensity landscapes called EEMs.[5] While the potential for accessing important information on biochemical processes via fluorescence is recognized,[6,7] a major obstacle is that the chemical and biological basis of fluorescence variability is poorly understood,[8] leading to difficulties in identifying appropriate monitoring parameters. A limited number of fluorescence peaks (referred to as Peaks A, B, C, M, and T) that apparently explained most differences between OM spectra were documented fifteen years ago.[5] However, even today, OM fluorescence is still most often described in broad terms as protein-like (tryptophan- or tyrosine-like), humic- or fulvic-like, reflecting its still-incomplete chemical characterization.

A significant advancement in the interpretation of EEMs occurred with the introduction to the field of Parallel Factor Analysis (PARAFAC).[9] PARAFAC enables the mathematical separation of chemically independent but spectrally overlapping fluorescence components. In theory, providing that the underlying assumptions of independence, linearity, and additivity are upheld, PARAFAC is capable of resolving true underlying chemical spectra.[10] In the case of natural organic matter; however, it is unclear how often this actually occurs, and despite a recent proliferation of PARAFAC studies of (D)OM data sets (e.g., refs 9 and 11−16), few authors have identified more than broad spectral similarities between PARAFAC models generated from unrelated data sets. Whether this may be a function of genuine differences between data sets, or is due to violations of the underlying model assumptions, different constraints or criteria imposed during modeling, or a combination of real differences and artifacts, remains to be elucidated. Furthermore, the chemical interpretations assigned to similar-appearing

PARAFAC components in different published studies is often contradictory, with "terrestrial" components in some models considered to be "marine" components in others, and "quinone" components indicative of redox state identified in some models but not in others.[9,11,13,16]

Clearly, a better understanding of organic matter fluorescence is needed, facilitating more coherent interpretations of PARAFAC components among studies. Furthermore, the strengths and limitations of PARAFAC in this application need careful examination. For example, we are aware of no studies in this field demonstrating that PARAFAC components provided greater resolution for discriminating among samples, or greater predictive power, than a well-chosen subset of EEM measurements. Further, a number of recent studies have refitted existing PARAFAC models to new data sets,[14,17,18] a practice that assumes that all important fluorophores in the new data set are accounted for in the original model, and that differences between the model and reality would not affect interpretation. Actual sensitivity analyses, however, are lacking.

These questions are particularly relevant to online monitoring applications, where any requirement to acquire, manipulate and interpret large data sets in real-time presents a significant cost and technological barrier. Continuous monitoring of organic matter fluorescence in oceans, rivers and water treatment facilities is of great interest,[19−22] because many parameters traditionally considered to be important indicators of water quality correlate with fluorescence, including total and dissolved organic carbon (TOC, DOC), Biochemical and Chemical Oxygen Demand (BOD, COD) and ultraviolet (UV) absorbance (reviewed in 23). However, inconsistent data processing [24] together with differences in wavelengths studied [23] obscure the identification of consistent trends across studies.

The aims of this study were to (1) obtain a robust characterization of fluorescence in a broad selection of Australian municipal water recycling treatment plants, using a diverse time-series data set incorporating multiple treatment stages, (2) identify consistent features across treatment plants to inform the design and selection of online monitoring tools, and (3) critically examine the behavior of PARAFAC when modeling large and diverse data sets in order to provide new insights on the strengths, limitations, and interpretation of PARAFAC models in the analysis of organic matter fluorescence.

■ METHODS

Six water recycling treatment plants in four Australian states were sampled over a 10−12 week period between March 2008 and June 2009, typically on a weekly schedule. These were Western Treatment Plant (Western TP, n = 246), Rouse Hill Recycled Water Plant (Rouse Hill RWP, n = 195), Sydney Olympic Park Water Reclamation and Management System (WRAMS, n = 364), St Mary'-s Replacement Flows Project Advanced Water Treatment Pilot Plant (St Mary's RFP, n = 261), Pimpama-Coomera Recycled Water Treatment Plant (Pimpama RWTP, n = 172), and Beenyup Pilot Reverse Osmosis Plant (Beenyup ROP, n = 241). Beenyup ROP and St Marys RFP were pilot scale operations that have since been replaced by full-scale plants operating at the same sites. Plant influents included municipal sewage, tertiary treated effluent and stormwater. Fifty-one sites were sampled representing various treatment stages; see the Supporting Information for details.

Whole water grab samples were collected in polyethylene terephthalate (PET) bottles or gamma-sterilized polypropylene (PP) bottles. Samples were 50 mL triplicates extracted in the laboratory from a 500 mL sample (Rouse Hill RWP) or 50 mL replicates collected by plant operators otherwise. Samples were transported cold to the University of New South Wales (UNSW) and analyzed within 48 h. Laboratory experiments indicated no fluorescent leachates from the PET or PP containers over this time frame.

No sample preparation was undertaken prior to spectral analysis. Absorbance spectra (200−600 nm) were measured in a 1 cm cell using a Cary Bio50 UV−vis spectrophotometer (Varian, Australia) and used to develop a matrix of correction factors for each fluorescence EEM accounting for inner filter effects.[25] Samples with absorbance exceeding 0.3 at 254 nm were excluded,[25] giving a data set of 1479 EEMs. Fluorescence was measured with a Varian Cary Eclipse spectrophotometer (Varian, Australia) for excitation wavelengths 200−400 nm in 5-nm intervals and emission wavelengths 280−500 nm in 2-nm intervals in a 1-cm cell. Different methods were applied for samples pre- and post-RO treatment. Pre-RO samples were scanned at 9600 nm min^{-1} with 5-nm slit widths on excitation and emission. Low fluorescence EEMs acquired from post-RO samples (14% of samples) were averages of three scans using 10-nm slit widths. EEMs were corrected for instrument bias and inner filtering effects, blank subtracted and normalized to the area under the water Raman peak of the blank at 350 nm, producing data in Raman Units.[24] Sample pH was measured in the laboratory with a HACH HQ14d portable pH meter (HACH, Australia); median pH in each plant ranged between 5.6 and 7.1 (see the Supporting Information).

Data Analysis. The fluorescence data sets were modeled using PARAFAC,[10] which uses an alternating least-squares algorithm to decompose the data signal into a set of trilinear terms and a residual array:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \quad i = 1,...,I; \; j = 1,...,J;$$

$$k = 1,...,K; \; f = 1,...,F$$

where x_{ijk} is the intensity of the ith sample at the jth variable (emission mode) and at the kth variable (excitation mode); a_{if} is directly proportional to the concentration of the fth analyte at emission wavelength j; b_{jf} is a scaled estimate of the emission spectrum of the fth analyte; c_{kf} is linearly proportional to the specific absorption coefficient (e.g., molar absorptivity) at excitation wavelength k and e_{ijk} is the residual noise, representing the variability not accounted for by the model.

PARAFAC models were generated using the PLS toolbox for Matlab (5.8, Eigenvector Research Inc.) with data scaled to unit variance in the sample mode and non-negativity constraints applied on all modes. Models were first derived for each treatment plant independently, and then for the global data set, producing seven models (Table 1). Measurements at excitation wavelengths below 240 nm excitation were excluded, as were outlier samples identified via influence and leverage plots (<1% of 1479 samples).[26] PARAFAC models were generated for 2−8 components, with the largest model selected for which spectra were reproduced identically in random halves of each data set (14 additional models). Scores were recovered by modeling the unscaled data using the validated excitation and emission loadings as preconditions, then multiplied by maximum excitation

Table 1. Description and Wavelength Positions of Peak Excitation (λ_{ex}) and Emission (λ_{em}) for PARAFAC Components in the Global Model (G) and Wavelengths That Track them in EEMs (E) with Correlation Coefficient R_{sq}[a]

	$\lambda_{ex}/\lambda_{em\ G}$	$\lambda_{ex}/\lambda_{em\ E}$	R_{sq}	description
G1	<250, 370/464	390/472	0.96	**H-ter:** Terrestrial humic-like fluorescence in high nutrient and wastewater impacted environments
G2	<250, 320/400	310/392	0.91	**H-mic:** Microbial humic-like fluorescence
G3	350/428	350/428	0.99	**WW:** Wastewater/nutrient enrichment tracer
G4	250/304	250/304	0.46	Protein - tyrosine-like, also observed in combination with G7
G5	<250/348	290/352	0.84	Protein, tryptophan-like, also observed in combination with G6
G6	290/352	290/352	1.0	**Tryp:** Protein, tryptophan-like
G7	270/300	270/304	0.97	**Tyro:** Protein - tyrosine-like

[a] See the Supporting Information for quantitative comparisons between G1−G7 and similar components in 11 previous studies.

and emission loadings to obtain maximum fluorescence of each component (F_{max}) in Raman units.

To compare between models, Tucker congruence coefficients[13] were determined for each excitation and emission spectrum from each plant model and compared to the global model. Two components were considered a match if they had the highest correlation coefficients on both excitation and emission spectra of all possible pairs of components.

To assess the potential for estimating PARAFAC scores from EEMs, the measured data were regressed against the maximum fluorescence of each component determined via PARAFAC. Thus for each wavelength pair in the EEM, an *n x 1* vector of concentrations were extracted from each data set of *n* EEMs and regressed against the *n x 1* vector of PARAFAC scores, with this procedure repeated for each component in the PARAFAC model. Linear regression was performed to obtain slope (m) and correlation coefficients (R^2) as a function of wavelength.

■ RESULTS AND DISCUSSION

Spectral Conformity among PARAFAC Models. PARAFAC models identified 5−7 independently varying fluorescing components in each plant. All models converged quickly and each was reproduced practically identically in independent halves of the data set. In the global model encompassing samples from all treatment plants, a total of seven PARAFAC components were identified (Figure 1), with the variance explained by each component in the model (of the normalized data set) decreasing sequentially for components G1−G7. When the model was applied to the unscaled data set in order to recover F_{max}, variance explained by each component in decreasing order was G3, G1, G5, G6, G2, G7, G4 with components G4 and G7 each accounting for <0.5% of total variability.

Comparisons of model spectra between plants revealed striking similarities (Figure 1). Components G1−3 representing humic- and fulvic-like fluorescence were identified in all six plants. All plants also had one tryptophan-like component (G6) as well as either tyrosine-like G4 and/or G7. Component G5, which had an emission spectrum almost identical to component G6, was found only in the global model; in individual models, component G5 was absorbed within the spectra of G6. Only two different components were identified exclusively in individual models (Figure 1, panel X8); one with an excitation spectrum very similar to G7 but with a blue-shifted emission spectrum was found in WRAMS and St Marys RFP (both in Sydney), and another with an excitation spectrum similar to C6 but a broader emission spectrum was unique to Beenyup ROP in Western Australia.
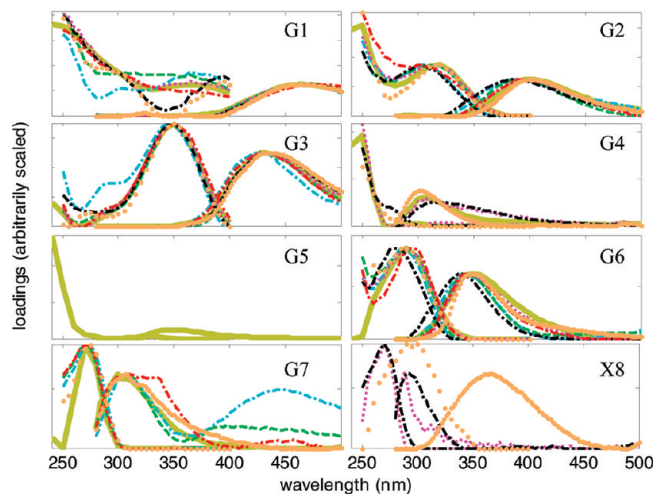

Figure 1. Comparison of PARAFAC components (excitation to the left of emission spectra) in samples from six water recycling plants. Spectra were derived from six independent models of fluorescence EEMs from Western TP, blue dash-dots; Rouse Hill RWP, green dashes; WRAMS, pink dots; St Marys RFP, black dash-dots; Pimpama RWTP, red dash-dots; Beenyup ROP, orange dots; and matched to similar components (G1-G7) in the global model of the entire data set (olive green solid line). Panel X8 depicts each unmatched component.

The spectra of components identified in different models varied in peak location and/or spectral shape. Differences in location were observed primarily for components G2 and G6, with spectral shifts up to ~15 nm observed in both excitation and emission spectra (Figure 1). The shape of the excitation spectrum was particularly variable for component G1 and for several other components at lower wavelengths (typically below 300 nm). Thus although component G1 exhibited a single emission spectrum, three slightly different excitation spectra were derived, each appearing in the model for two plants.

The irregular excitation spectra for component G1 in the Western TP and Rouse Hill RWP models coincided with a spurious secondary emission peak for component G7 having similar characteristics to the G1 emission spectra, and in the Rouse Hill RWP model, slight distortion of component G3. This indicates that at times when PARAFAC had difficulty distinguishing two components, the models returned hybridized spectra. Spectra for the global model, although not conforming to an average spectrum, lay within the range of spectra exhibited across the individual plants.

Although the convergence of all seven models upon similar spectral shapes provides convincing evidence that real chemical

phenomena were identified, core consistency (CC) statistics that track the adherence of the model to the PARAFAC algorithm were well below the ideal of 100%, indicating possible overfitting.[27] No six-component model could be validated in split halves. However, a five-component global model that combined component G5 with G6 and component G4 with G7 had a relatively high core consistency (CC = 60%). The cleaving of components G4 and G7 was observed in several individual models (Figure 1) but the cleaving of components G5 and G6 appears to be due to the X8 (Beenyup) component being absorbed into the G6 spectra, and is probably erroneous. Modeling by treatment stages provided further support for a six component solution. Independent PARAFAC models of the input (secondary and tertiary treated effluent, $n = 196$) and output (recycled water, $n = 158$) streams identified six components with similar spectral shapes. This indicates that no new components were generated within the plants and all were typically detectable following treatment.

Table 1 summarizes the peak characteristics of each component in the global model. Comparisons with spectral data obtained for 11 published PARAFAC models plus reference compounds, made by overlaying spectra and calculating factor similarity statistics, are provided in the Supporting Information. These show that each PARAFAC component identified in this study is common to at least two previous models of fluorescence in wastewater, ballast water, lakes, estuaries or the ocean. Except for shifts in excitation maxima, the spectra for three components were almost identical to spectra for the pure compounds syringaldehyde (C3), tryptophan (G6) and tyrosine (G7).

Selective Removal of Components through Treatment. The efficiency of the water recycling treatment process in removing individual fluorescent components is illustrated in Figure 2. For each plant, intensities are shown for the five components common to all individual models, identified as "WW" (G3 in the global model), "H-ter" (terrestrial humic G1), "H-mic" (microbial humic G2), "Trypt." (tryptophan-like G6), and "Tyros." (tyrosine-like G7).

In the three plants that did not implement reverse osmosis (Figure 2A–C), most removal occurred in the first or last treatment stages, coinciding with initial chlorine dosing. At Pimpama RWTP, chlorination designed to achieve 10–15 mg L^{-1} residual free Cl was implemented prior to the dual media filtration process. Strong removal between stages 1–2 compared to negligible removal by later UV and chlorination (between stages 3–5) indicates that all of the easily oxidizable organics were already oxidized by stage 2 (Figure 2Bii). At Rouse Hill RWP, a large effect was associated was chlorination implemented after UV treatment (concentration × time, CT >300 mg L^{-1} min^{-1}), whereas in Western TP (CT > 30 mg L^{-1} min^{-1}), the effect was smaller. Overall, "WW" component (G3) was removed most effectively in all three plants, particularly by chlorination, whereas removal of tyrosine-like fluorescence was relatively poor. The fluorescent signature of the recycled product varied between plants, but was typically dominated by terrestrial and microbial humic fluorescence (Figure 2A–Ciii). In the remaining three plants (see the Supporting Information), nearly complete removal of organics occurred following reverse osmosis.

Identification of F_{max} from EEMs. The principle of parsimony indicates that complex models are unjustified unless they robustly deliver better solutions than simpler models. Therefore, the relationship between PARAFAC scores and EEM measurements, as would be obtained from simple "peak peaking", was examined. Figure 3 shows correlations between PARAFAC intensities (F_{max}) for components W1–W5 in the Western TP model and fluorescence intensities in the original EEMs. Linear regression fits (R^2) range between 0.01 to >0.99 across the EEMs (left-hand panels), with the strongest correlations occurring near the PARAFAC component peaks (depicted by white crosses), and lower correlations elsewhere. Regression slopes in the range 0–1.5 are shown in right-hand panels, with slopes equal to 1.0 indicating that F_{max} was identical (excluding random error) to intensity in the EEM.

For component W1 in Figure 3, the regression fit and slope (m) at the position of the secondary excitation peak ($\lambda_{ex}/\lambda_{em} = 300/390$ nm) were both close to 1.0, indicating that F_{max} for component W1 was equivalent to EEM intensity at wavelength position $\lambda_{ex}/\lambda_{em} \sim 300/390$ nm. For similar reasons, EEMs gave accurate estimates of F_{max} for components W2, W4 and W5. For component W3, the maximum slope was ~ 0.6 due to the emission overlap with component W2, thus W3 can be only be accurately estimated from EEMs provided that the relative proportions of W2 and W3 do not change significantly. Component W1, W3, and W4 peaks are located in EEM regions where slopes were relatively stable, indicating that small shifts in the wavelength position of measurement would not significantly alter the equation for estimating F_{max} of these components. Conversely, the W2 and W5 peaks are located in regions where m changes rapidly, indicating that the equation is sensitive to precise measurement location. These results indicate that monitoring raw fluorescence at a small number of appropriate wavelengths could capture the same essential information as would PARAFAC analyses of full EEMs.

Sensitivity of F_{max} to Model Inadequacies. An important criterion for any complex data manipulation is that the derivation and interpretation of models is both unambiguous and robust to small perturbations in the sample set or measurement parameters and to variations in equipment or operators. A potential weakness of PARAFAC is that due to the interdependence of the simultaneously estimated components in the model, the inclusion of one or more poorly estimated components significantly affects the spectra and scores of all other components. The effect of applying imperfect PARAFAC models is illustrated clearly by comparisons of component intensities according to individual models versus the global model (Table 2, see also the Supporting Information). We consider alternative PARAFAC models to be interchangeable only if both of the following conditions are fulfilled: (1) Intensities (F_{max}) for corresponding components are linearly correlated with high R^2, and (2) slope $m = 1$, implying that any differences in spectra between models do not significantly affect estimated intensities.

In paired comparisons of corresponding individual and global components, the first condition was nearly always fulfilled (typically $R^2 > 0.95$), however, regressions slopes ranged between 0.73–1.22 and were almost always significantly different from 1.0 (Table 2). Thus, significantly different fluorescence intensities and ratios of components were predicted using each individual model compared to the global model. This result highlights the problems inherent in relying upon single PARAFAC representations of diverse sample sets. It also implies that the practice of forcing existing PARAFAC models onto new fluorescence data sets is very likely to generate systematically biased estimates of fluorescence intensities and fluorescence ratios if the refitted model does not completely and accurately
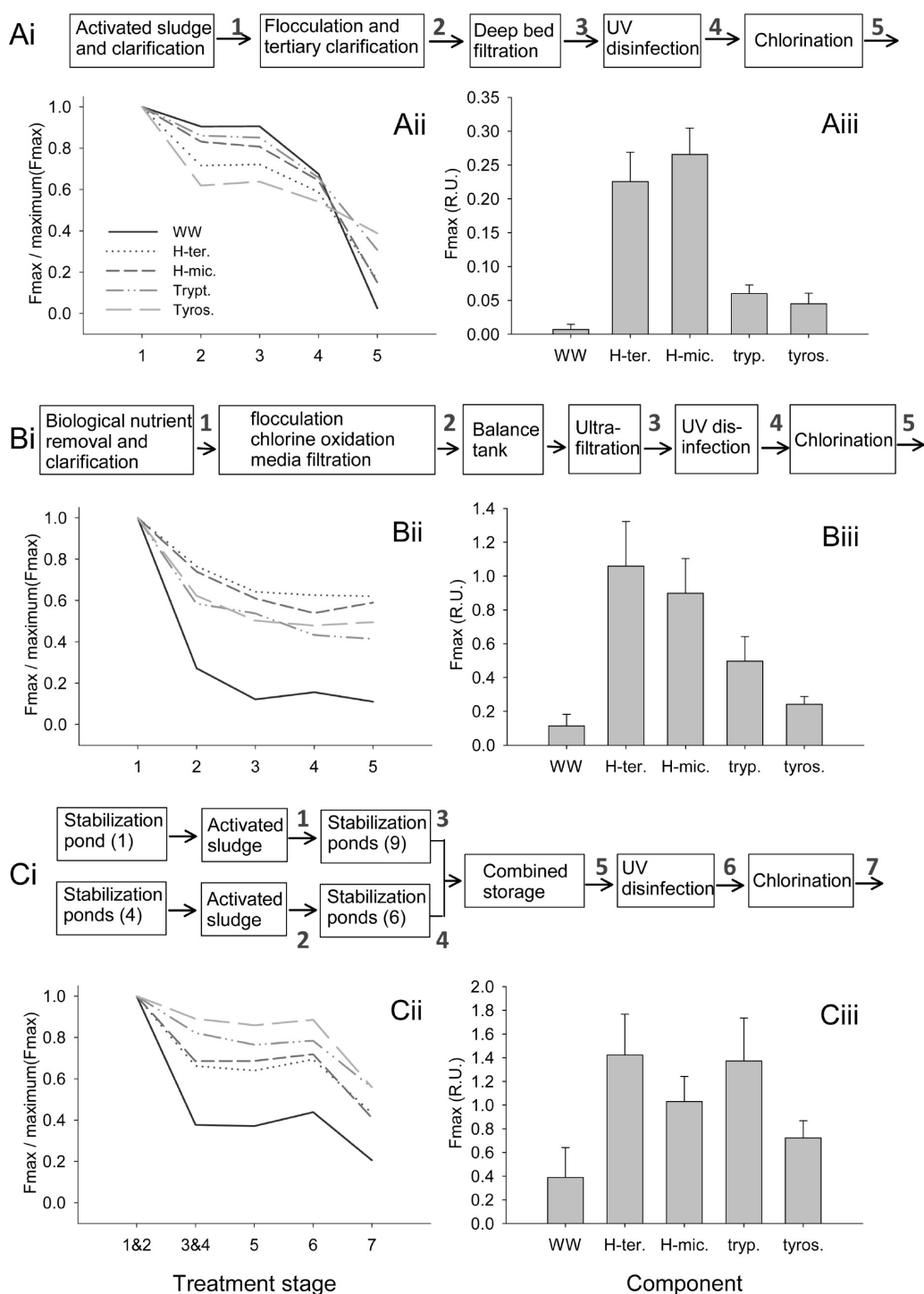
Figure 2. Removal of organic matter through the water treatment process at (A) Rouse Hill RWP, (B) Pimpama RWTP, and (C) Western TP. Subplots show (i) the process train; (ii) persistence of PARAFAC components through sequential treatment stages (mean of F_{max} at each process stage divided by maximum F_{max} across all stages); and (iii) final intensities of components following treatment (mean + SD).

represent spectra for each of the underlying components in the new data set.

Toward a Unified PARAFAC Model. We are aware of more than 30 published PARAFAC models of OM fluorescence in natural waters, together with a growing number from waste-waters and industrial processes. While each new study draws attention to similarities with previously identified spectra, notice-able differences are just now beginning to emerge. Component

G1 in this study, which was the dominant terrestrial humic-like fluorescence, is superficially similar to at least one component found in nearly every PARAFAC model of OM fluorescence. However, a precise match for its emission spectrum was observed only in three models that included samples from engineered, wastewater impacted environments [11,15,28] (see the Supporting Information) and in marine mesocosms installed along a Danish fjord and subject to nitrogen and phosphorus enrichment.[7]
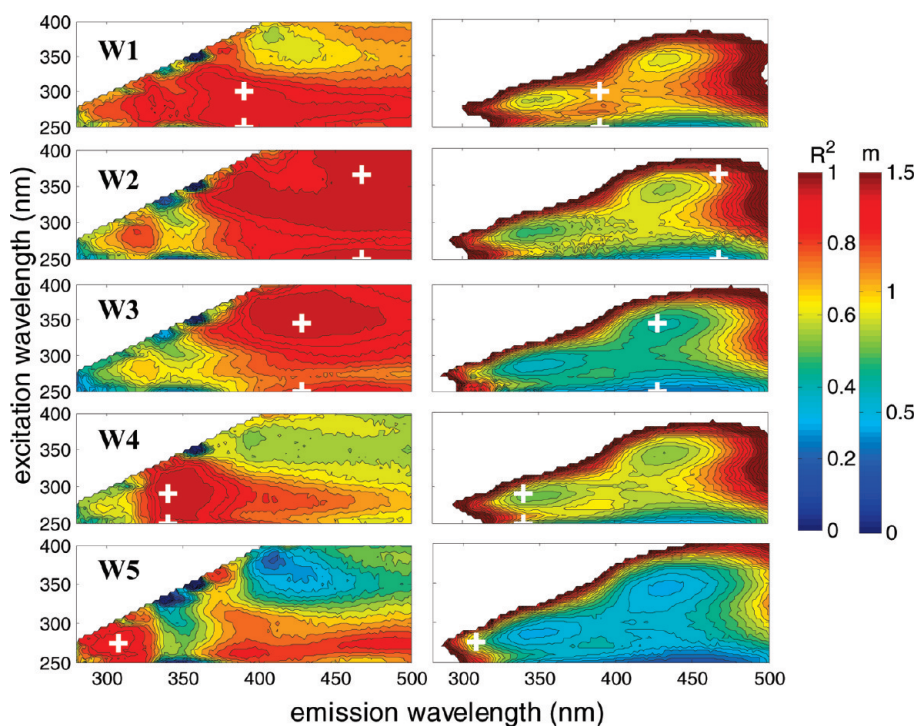
Figure 3. Recovery of F_{max} from fluorescence EEMs via linear regression through the origin, demonstrated for Western TP. Plots show R^2 on left and regression coefficients (m) on right for each component (W1–W5) in the individual PARAFAC model. White crosses locate the excitation and emission peaks of the PARAFAC component.

Table 2. 95% Confidence Intervals for the Slope of the Linear Regression of F_{max} in Individual Models versus the Global Model for PARAFAC Components G1, G2, G3, and G6[a]

		component			
plant	n	G1	G2	G3	G6
Beenyup ROP	241	**0.85–0.85**	**0.85–0.85**	**0.91–0.91**	**0.80–0.82**
Pimpama RWTP	172	**0.99–0.99**	**1.19–1.22**	**1.16–1.20**	**0.85–0.88**
Rouse Hill RWP	195	**1.18–1.18**	1.00–1.00	**0.94–0.94**	**0.88–0.90**
WRAMS	364	**0.93–0.94**	**0.97–0.98**	**1.04–1.04**	
St Marys RFP	261	**0.85–0.86**	**1.08–1.19**	**1.16–1.17**	0.95–1.00
Western TP	246	**0.73–0.74**	**1.04–1.07**	**1.17–1.19**	**0.85–0.86**

[a] Intervals excluding 1.00 (**bold** type) indicate significantly different intensities according to the two models, blank cell indicates no correlation.

Nearly all published models also include one or more microbial ("M") fluorescence peaks similar to G2, e.g., refs 7, 11, and 13. Component G3 appears to have been observed twice previously, in North American wastewaters [15] and in the Danish study,[7] where it was produced by microbial degradation. Like G1, G3 appears to be a fluorescence component found in nutrient-rich environments.

Similarly, tryptophan- and tyrosine-like components are common to practically all PARAFAC models with four or more components, e.g., refs 9 and 11–15; however, in our study these both had "shadow components" with very similar spectral characteristics. Thus, two of the three "unmatched" spectra (X8 in Figure 1) were nearly identical to PF6 described by Dixon[12] from wastewaters and laundry effluent in Florida and caused by detergents and optical brighteners. In the current study, PF6 appeared in two individual models, whereas in the other four, it was indistinguishable from tyrosine.

Insights are offered into strengths and limitations of PARAFAC applied to OM fluorescence spectra. In this fifteen-month study of six Australian water recycling treatment plants, PARAFAC models for different plants were surprisingly similar, despite geographically and temporally variable plant influents, variable pH and slit widths and a data set comprising of nearly 1500 samples from 51 separate water treatment stages. This indicates that existing PARAFAC models can help to predict the number and characteristics of spectral components likely to be present in new environments involving similar types of sources, which has implications for all aspects of experimental design, from instrumentation through to sample size selection. In carefully designed studies, PARAFAC could be a powerful ally for examining spectral shifts due to pH, interactions with metal ions, and other matrix effects with greater sensitivity than has been possible from studying total fluorescence in EEMs. In this study, explicit testing of the effects of pH on spectral shape was precluded because pH and treatment process stage were confounded variables.

Broad similarities between OM fluorescence spectra across world regions would seem to imply that across similar types of environments, the number of spectrally unique fluorophores contributing to natural organic matter fluorescence may be relatively few, such that populating a global database of OM PARAFAC spectra could represent a useful and achievable goal. At the same time, limitations to the chemical interpretation of PARAFAC components have not received enough attention in the literature. In this study, the act of combining six data sets did not increase the number of PARAFAC components identified. Instead, it caused shifts in the positions of some components and modified the spectra of others. This foreshadows limits to the sensitivity of fluorescence for distinguishing individual organic

matter sources in water mixtures. Further, the presence of a tryptophan-like "shadow component" in one plant (Beenyup ROP, component X8) caused the erroneous splitting of components G5 and G6 in the global model. Clearly, PARAFAC spectra from OM data sets describe the probabilistic distribution of an ensemble of individual spectra belonging to a range of spectrally similar chemical moieties from a range of sources, rather than exact chemical spectra. Known sources of variability in this study (e.g., pH, treatment processes) would undoubtedly have contributed to PARAFAC's difficulty. At the same time, diverse data sets are very likely to be subject to known or unknown sources of spectral variability,[8,24] especially when samples encompass a wide range of environmental conditions, e.g., refs 4 and 16. The effect of such diversity on the development and application of "universal" PARAFAC models deserves wider recognition.

Implications for Online Monitoring. Several results from this study have direct implications for online fluorescence monitoring of water treatment plants, including (1) a limited number of spectrally similar PARAFAC components explained nearly all of the OM fluorescence variability in water recycling treatment plants across Australia; (2) similar components were found in independent models of initial and final treatment stages, indicating that these components were neither generated within the plants nor typically entirely removed; (3) PARAFAC intensities (F_{max}) could be estimated accurately by multiplying directly measured fluorescence by predetermined regression coefficients for each component; (4) the magnitudes of regression coefficients varied in their sensitivity to deviations in excitation or emission wavelength.

These results imply that instrumentation for online fluorescence monitoring need not be specific to plant or process stage. In the regression of PARAFAC scores to unprocessed EEMs (Figure 3), regression coefficients near 1.0 for many wavelength pairs highlight the autocorrelation of large amounts of data in EEMs. Thus across all plants and process stages, the scores for components G1-G3 and G6-G7 could be estimated quite accurately ($R^2 = 0.91-1.00$) from measurements at the discrete excitation/emission wavelengths listed in Table 1. This indicates that monitoring via a small number of simple fluorometers with appropriate wavelength selectivity should capture essentially the same information as would online monitoring of full EEMs. By default, filter-based in situ fluorometers collect fluorescence signals over a range of wavelengths, which should ameliorate the effect of slight variations in either the optical parameters of the instrument or the fluorescence characteristics of samples. However, Figure 3 suggests that greater precision and smaller bandpasses may be more important for monitoring some fluorescent components than others.

ASSOCIATED CONTENT

Ⓢ Supporting Information. Nine appendices are provided reporting sample characteristics, OM removal through treatment in RO plants, PARAFAC model loadings, model validations, comparisons with previous studies, and sensitivity analyses (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: krm@unsw.edu.au.

REFERENCES

(1) Stedmon, C. A.; Markager, S.; Søndergaard, M.; Vang, T.; Laubel, A.; Borch, N. H.; Windelin, A. Dissolved organic matter (DOM) export to a temperate estuary: seasonal variations and implications of land use. *Estuarine Coastal* **2006**, *29* (3), 388–400.

(2) Coble, P. G.; Green, S. A.; Blough, N. V.; Gagosian, R. B. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. *Nature* **1990**, *348*, 432–435.

(3) Ahmad, S. R.; Reynolds, D. M. Monitoring of water quality using fluorescence technique: Prospect of on-line process control. *Wat. Res.* **1999**, *33* (9), 2069–2074.

(4) Murphy, K. R.; Ruiz, G. M.; Dunsmuir, W. T. M.; Waite, T. D. Optimized parameters for fluorescence-based verification of ballast water exchange by ships. *Environ. Sci. Technol.* **2006**, *40* (7), 2357–2362.

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

(6) McKnight, D. M.; Boyer, E. W.; Westerhoff, P. K.; Doran, P. T.; Kulbe, T.; Andersen, D. T. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **2001**, *46* (1), 38–48.

(7) Stedmon, C. A.; Markager, S. Tracing the production and degradation of autochthonous fractions of dissolved organic matter using fluorescence analysis. *Limnol. Oceanogr.* **2005**, *50* (5), 1415–1426.

(8) Boyle, E. S.; Guerriero, N.; Thiallet, A.; Vecchio, R. D.; Blough, N. V. Optical Properties of Humic Substances and CDOM: Relation to Structure. *Environ. Sci. Technol.* **2009**, *43* (7), 2262–2268.

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

(10) Bro, R. PARAFAC: Tutorial and applications. *Chemom. Intell. Lab.* **1997**, *38*, 149–171.

(11) Chen, M. L.; Price, R. M.; Yamashita, Y.; Jaffe, R. Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. *Appl. Geochem.* **2010**, *25* (6), 872–880.

(12) Dixon, L. K. *Optical Brighteners: PARAFAC Analyses of EEM Fluorescence Data for the Conceptual Design of Field Instrumentation and Methods*; Mote Marine Laboratory Technical Report No. 1316; Mote Marine Laboratory: Sarasota, FL, 2009.

(13) Murphy, K. R.; Stedmon, C. A.; Waite, T. D.; Ruiz, G. M. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Mar. Chem.* **2008**, *108* (1–2), 40–58.

(14) Yamashita, Y.; Jaffé, R. Characterizing the Interactions between Trace Metals and Dissolved Organic Matter Using Excitation-Emission Matrix and Parallel Factor Analysis. *Environ. Sci. Technol.* **2008**, *42* (19), 7374–7379.

(15) Yu, G.-H.; He, P.-J.; Shao, L.-M. Novel insights into sludge dewaterability by fluorescence excitation-emission matrix combined with parallel factor analysis. *Wat. Res.* **2010**, *44* (3), 797–806.

(16) Cory, R. M.; McKnight, D. M. Fluorescence Spectroscopy Reveals Ubiquitous Presence of Oxidized and Reduced Quinones in Dissolved Organic Matter. *Environ. Sci. Technol.* **2005**, *39* (21), 8142–8149.

(17) Fellman, J. B.; Miller, M. P.; Cory, R. M.; D'Amore, D. V.; White, D. Characterizing Dissolved Organic Matter Using PARAFAC Modeling of Fluorescence Spectroscopy: A Comparison of Two Models. *Environ. Sci. Technol.* **2009**, *43* (16), 6228–6234.

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

(19) Carstea, E. M.; Baker, A.; Bieroza, M.; Reynolds, D. Continuous fluorescence excitation-emission matrix monitoring of river organic matter. *Wat. Res.* **2010**, *44* (18), 5356–66.

(20) Conmy, R. N.; Coble, P. G.; Castillo, C. E. D. Calibration and performance of a new in situ multi-channel fluorometer for measurement of colored dissolved organic matter in the ocean. *Cont. Shelf Res.* **2004**, *24* (3), 431–442.

(21) Downing, B. D.; Boss, E.; Bergamaschi, B. A.; Fleck, J. A.; Lionberger, M. A.; Ganju, N. K.; Schoellhamer, D. H.; Fujii, R. Quantifying fluxes and characterizing compositional changes of dissolved organic matter in aquatic systems in situ using combined acoustic and optical measurements. *Limnol. Oceanogr. Meth.* **2009**, *7*, 119–131.

(22) Spencer, R. G. M.; Pellerin, B. A.; Bergamaschi, B. A.; Downing, B. D.; Kraus, T. E. C.; Smart, D. R.; Dahlgren, R. A.; Hernes, P. J. Diurnal variability in riverine dissolved organic matter composition determined by in situ optical measurement in the San Joaquin River (California, USA). *Hydrol. Process.* **2007**, *21* (23), 3181–3189.

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

(24) Murphy, K. R.; Butler, K. D.; Spencer, R. G. M.; Stedmon, C. A.; Boehme, J. R.; Aiken, G. R. The measurement of dissolved organic matter fluorescence in aquatic environments: An interlaboratory comparison. *Environ. Sci. Technol.* **2010**, *44*, 9405–9412.

(25) Ohno, T. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. *Environ. Sci. Technol.* **2002**, *36*, 742–746.

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

(27) Bro, R.; Kiers, H. A. L. A new efficient method for determining the number of components in PARAFAC models. *J. Chemometr.* **2003**, *17* (5), 274–286.

(28) Zhuo, J. F.; Guo, W. D.; Deng, X.; Zhang, Z. Y.; Xu, J.; Huang, L. F. Fluorescence Excitation-Emission Matrix Spectroscopy of CDOM from Yundang Lagoon and Its Indication for Organic Pollution. *Spectrosc. Spectr. Anal.* **2010**, *30* (6), 1539–1544.