Continuous fluorescence excitation–emission matrix monitoring of river organic matter

Elfrida M. Carstea [a,*], Andy Baker [b], Magdalena Bieroza [c], Darren Reynolds [d]

[a] National Institute of R&D for Optoelectronics, INOE 2000, 409 Atomistilor Str., RO-077125 Magurele, Romania
[b] Connected Waters Initiative, University of New South Wales, Water Research Laboratory, 110 King St, Manly Vale, NSW 2093, Australia
[c] Department of Civil Engineering, Queen's Building, University Walk, Bristol BS8 1TR, UK
[d] University of the West of England, Coldharbour Lane, Bristol BS16 1QY, UK

ARTICLE INFO

ABSTRACT

Real-time fluorescence monitoring has been mostly performed in marine systems, with little progress being made in the application of fluorescence excitation–emission matrix (EEM) spectroscopy, especially for freshwater monitoring. This paper presents a two weeks experiment where real-time fluorescence EEM data have been obtained for Bourn Brook, Birmingham, UK, using an in-situ fibre-optic probe. Fluorescence EEMs were measured every 3 min for two weeks, with control 'grab' samples every hour analyzed for fluorescence EEMs as well as pH, conductivity and dissolved organic carbon. Comparison of real-time and control samples showed an excellent agreement, with no evidence of fibre-optic probe fouling. EEMs of different character were identified using self-organizing maps, which demonstrated seven clusters of fluorescence EEMs which related to the intensity of fluorescence and relative intensities of peak T_1 and T_2 vs. peak C and peak A fluorescence. Fluorescence intensity of peaks A and C were observed to increase with rainfall, and a diesel pollution event was detected through an increase in T_2 fluorescence.

1. Introduction

Fluorescence spectroscopy has been intensely used in recent decades for the analysis of dissolved organic matter (DOM) in water. The fluorescent DOM fractions, which are typically observed in fresh waters and which provide water quality information are tryptophan-like and tyrosine-like, indicators of microbial activity and humic-like, derived from both allochthonous and autochthonous sources. The excitation/emission wavelength pairs are specific to each component and are A and C for humic substances ($\lambda_{\text{excitation}}/\lambda_{\text{emission}}$ $\sim 225/400$–500 nm corresponding to A, 300–350/400–500 nm corresponding to C) and T_1 and T_2 for tryptophan ($\lambda_{\text{excitation}}/$ $\lambda_{\text{emission}}$ $\sim 280/\sim 350$ nm – T_1, $\sim 225/\sim 350$ nm – T_2), B for tyrosine ($\lambda_{\text{excitation}}/\lambda_{\text{emission}}$ $\sim 225/\sim 305$ nm). These fluorophores are reviewed, in more detail, in Hudson et al. (2007).

Many studies have shown that this technique can characterize natural organic matter (Baker and Spencer, 2004; Winter et al., 2007) and detect different types of aquatic pollutants, like sewage (Baker, 2001; Reynolds, 2002), oil (Bugden et al., 2008; Patra and Mishra, 2002) or pesticides (JiJi et al., 1999, 2000). Furthermore, DOM fluorescence data correlate with water quality parameters such as total organic carbon (Vodacek et al., 1995), aquatic plankton (Mopper and Schultz, 1993) and biological oxygen demand (Reynolds and Ahmad, 1997; Hudson et al., 2008). Based on fluorescence

spectroscopy capabilities, Ahmad and Reynolds (1999) have suggested applying this method for on-line process control in sewage treatment plants. Subsequently, other studies have implied the use of fluorescence spectroscopy as a potential monitoring tool for recycled water (Henderson et al., 2009), drinking water treatment processes (Cheng et al., 2004; Bieroza et al., 2009a) and urban watersheds with sewage effluents (Hur et al., 2008). Despite these implications, few continuous real-time fluorescence experiments have been performed on freshwater, mostly due to the lack of proper instrumentation.

Until now, real-time fluorescence data have mostly been obtained for seawater with instruments specifically designed for this application (Chen, 1999; Barbini et al., 2003; Conmy et al., 2004; Drozdowska, 2007). Similar equipment has been used for freshwater studies in order to obtain continuous real-time data (Spencer et al., 2007; Downing et al., 2009). However, the continuous DOM monitoring was limited by the fixed excitation and emission wavelengths, allowing the measurement of only the terrestrial components. Another limitation was the clogging of pre-filters which required frequent filter replacement. Continuous monitoring has also been performed by Carstea et al. (2009) who collected water samples at hourly scale, but the measurements were made within 24 h and not in real-time. These studies revealed that dissolved organic matter varied at a daily timescale, depending on the river type, and was highly influenced by precipitation. However, this study raised many topics for further research such as: does dissolved organic matter exhibit variability at a higher frequency; is there different behavior in the protein-like and humic-like components; if and for how long can on-line measurements be performed without any cleaning procedure; whether there are interferences in the fluorescence signal from biofilm formation; what is the optimum measurement frequency? To attempt to answer these points, here we continuously monitor, in-situ and real-time, the water quality of a small urban catchment using a standard bench-top fluorometer and fibre-optic probe. Also, the fluorescence properties of a diesel pollutant, which was been detected during the real-time monitoring, is presented and analyzed using both peak-picking techniques and the expert system approach of self-organizing map (SOM). SOM enables robust pattern recognition of continuous fluorescence data based on the information from the entire EEMs and not only on arbitrary selected peak values.

SOMs are unsupervised neural network algorithms, which represent an excellent tool for fluorescence data analysis (Rhee et al., 2005; Bieroza et al., 2009a). Usually, EEM data analysis is done with other multiway techniques, Principal Component Analysis (PCA) and Parallel Factor Analysis (PARAFAC), but SOM offers some important advantages over these methods: high tolerance to noise and less time-consuming interpretation and validation of final components (Bieroza et al., 2009a). The SOM algorithm explores the input data to find and extract fluorescence features, describing an elementary pattern of information that represents the presence and spectral properties of particular fluorophores or groups of fluorophores (Kohonen, 2001). The fluorescence feature extraction involves nonlinear transformation of the EEMs onto a two-dimensional map. Samples containing fluorophores of similar quantitative (intensity) and qualitative (emission and excitation wavelengths) properties are clustered in the same regions of the SOM. When applied to analysis of continuous fluorescence data like in this study, a SOM-based expert system can provide on-line evaluation of fluorescence properties and detection of diesel pollution.

Recently, a robust SOM technique was successfully used to examine the fluorescence data characterizing drinking water quality and to reveal spatial and temporal patterns in organic matter and fluorescence properties (Bieroza et al., 2009b). Aymerich et al. (2009) showed that SOM method can effectively classify phytoplankton fluorescence spectra and Rhee et al. (2005) could discriminate between fermentation processes using SOM analysis of 2D fluorescence spectra. A more detailed description of the SOM algorithm can be found in Kohonen (2001).

2. Materials and methods

2.1. Site description

Measurements were performed on a small urban river, the Bourn Brook, from Birmingham, UK, which has a catchment area of 27.91 km^2 (52°26.51 N; 1°55.49 W; Fig. 1). Apart from the natural catchment sources, the river receives water only from storm sewer systems and combined sewer overflows. The UK Environment Agency most recently classified the river chemical water quality as "good" (grade B) and biological water quality as "fair" (grade D). During the period 2005 and 2007, Environment Agency routine chemical water quality monitoring yielded the following mean values: ammonia 0.11 mgN/l, dissolved oxygen 91.3%, nitrate 9.6 mg/l, phosphate 0.12 mg/l. Biological water quality using invertebrates scored an observed 11 taxa (against an expected 24.7) with an average score per taxa of 4.27 (against an expected score of 5.73).

2.2. Experimental design

The experiment was performed on 4–14 August 2009. Fluorescence EEMs were recorded every 3 min using a Varian Cary Eclipse spectrofluorometer, housed in a mobile laboratory that was left bank-side to the river. The spectrofluorometer was equipped with a fibre-optic probe and 1 cm path length probe tip. The probe was inserted into a 50 mL sample chamber, into which river water was continuously pumped at a flow rate of ~30 ml/min. Pre-filtering occurred in the river with three parallel filter screens: a course mesh to screen for course debris and then filters of 2 and 0.5 mm pore size. In order to test the reliability of this design, no cleaning was performed of any part of the apparatus. No clogging of the filters occurred, although significant discolouration was observed along with some fine sediment accumulation on the outer filter. Biofilm fragments were visually observed in the tubing and sample chamber, continuously increasing over the experimental period. The system ran for 11 days until a power supply failure occurred, with some small gaps in the data due to software crashes.

Fig. 1 − Sampling site on the Bourn Brook catchment.

Fluorescence spectra, in the form of excitation−emission matrix (EEM), were recorded in the excitation wavelength range 225−400 nm and emission wavelength range 280−500 nm, with scan rate of 9600 nm/min. Bandpass was set at 10 nm and 5 nm for excitation and emission, respectively, the former wider than conventionally used on our instrument to account for the lower optical efficiency of the fibre-optic system. In case fine turbidity affected the fluorescence, an emission filter between 295 and 1100 nm was used to remove any scattered light. Instrument stability was tested daily by inserting the fibre-optic probe into a bottle of deionised water to measure the Raman peak at 348 nm excitation wavelength and 395 nm emission wavelength. Raman values were almost constant throughout the experiment, having an average value of ∼7 a.u. and standard deviation of ∼1.5 a.u. The fluorescence intensity of all spectra was corrected to the average Raman value of 7 a.u. and normalized to a maximum value of 1000 a.u. Apart from this, no other excitation or emission corrections were applied.

Peak-picking method implied the use of Excel macros (Office 2003), whereas SOM calibration was carried out in Matlab 7.7 with the Statistics Toolbox 7.0 and Neural Network Toolbox 6.0.1, generating a map of size 20 by 13 units. SOM training was carried out based on a procedure described elsewhere (Bieroza et al., 2009b). In essence, SOM training is an iterative process in which, for each input sample comprising an unfolded EEM, the neuron with reference vector weights most similar to the input vector is first identified (winner or best-matching unit, BMU). Thus, for each input EEM presented to the SOM network, the output neuron with the reference vector most similar to the vector representation of EEM is

selected. Once the best-matching reference vector for each input EEM vector is found, its weights and the weights of its neighbouring neurons are modified and moved towards the input vector (self-organization feature of the algorithm) (Eq. (1)):

$$w_i(k + 1) = w_i(k) + \epsilon(k)h_p(i, k)\{x_j(k) - w_i(k)\} \tag{1}$$

where: $w_i(k)$ is the previous weight of neuron, $w_i(k + 1)$ is the new weight of neuron, $\epsilon(k)$ is the learning rate, $h_p(i, k)$ describes the neighbourhood of the winning neuron, k is the number of epochs (a finite set of input patterns presented sequentially) and p is the index of the winning neuron. The learning rate describes the speed of the training process $(0 < \epsilon(k) < 1)$ and decreases monotonically during the training phase. The topological neighbourhood can be described as a neighbourhood set of array points, Nc, around the given node c. During the training of the map, the radius of the Nc (a size of the neighbourhood set) decreases monotonically to enable the global ordering of the map. Thus, the projection of the input data is done in two phases: the main training (large Nc radius) and fine adjustment of the map (small Nc radius) (Kohonen, 2001).

The trained network activates the appropriate output neurons of the network according to the input samples without prior knowledge of the process that produced the data and its distribution. Consequently, the analysis of the networks' output provides the basis for extraction of relationships and regularities from the original data.

Conductivity and pH were measured using a Myron meter. Also, several samples were tested for DOC with Shimadzu TOC − Vcpn analyzer. Samples selected for DOC measurements were previously filtered and acidified to pH 2, by adding

**Table 1 – Daily averaged values for fluorescence spectroscopy results and standard parameters.**

Date	Peak T₁ (a.u.)[a]	Peak T₂ (a.u.)	Peak C (a.u.)	Peak A (a.u.)	Peak T₁ Em λ (nm)	Peak T₂ Em λ (nm)	Peak C Em λ (nm)	Peak A Em λ (nm)	Peak T₁ (GS[b]) (a.u.)	Peak T₂ (GS) (a.u.)	Peak C (GS) (a.u.)	Peak A (GS) (a.u.)	pH	Conductivity (μS/cm)	DOC (mg/L)
4/8/2009	76	31	62	121	358	362	429	431	82	32	68	141	7.8	338	4
SD[c]	9	3	7	17	6	2	11	13	5	4	8	13	0.2	20	-
5/8/2009	68	29	52	118	357	360	430	429	77	30	63	162	7.7	403	4
SD	9	2	4	16	6	4	13	13	9	3	8	26	0.4	26	0.3
6/8/2009	56	25	45	101	355	357	428	429	61	25	54	127	7.6	504	3
SD	6	2	4	14	6	4	13	13	9	3	8	26	0.2	7	0.2
7/8/2009	75	33	50	117	354	357	432	431	74	33	56	141	7.6	379	4
SD	8	2	4	14	6	4	14	15	4	2	6	10	0.2	19	1
8/8/2009	53	24	40	90	356	358	434	434	51	22	45	123	7.7	496	3
SD	7	2	4	14	6	5	15	14	5	2	4	38	0.1	3	0.5
9/8/2009	46	20	36	83	355	357	429	434	48	22	44	106	7.7	485	2
SD	5	2	3	12	6	5	14	154	6	2	6	9	0.1	3	0.2
10/8/2009	152	43	35.	83	352	354	427	432	142	40	42	99	7.7	487	3
SD	188	42	3	9	7	7	11	15	271	53	6	13	0.1	6	0.8
11/8/2009	324	76	37	94	343	345	429	429	847	180	44	115	7.6	487	4
SD	54	11	8	22	2	4	14	15	46	10	3	24	0.3	7	1.9
12/8/2009	202	68	51	117	346	354	432	428	363	88	56	143	7.5	299	5
SD	41	16	12	23	4	6	15	12	140	29	12	37	0.2	68	2.3
13/8/2009	149	47	39	99	346	350	435	434	230	60	43	124	7.9	452	3
SD	16	6	5	16	4	6	17	16	20	5	5	13	0.2	12	0.9
14/8/2009	134	49	34	83	346	348	436	434	176	53	39	111	7.8	478	2
SD	11	6	8	11	5	5	17	16	10	5	4	15	0.2	9	0.6

a a.u. – arbitrary units.
b GS – grab sample.
c SD – standard deviation.
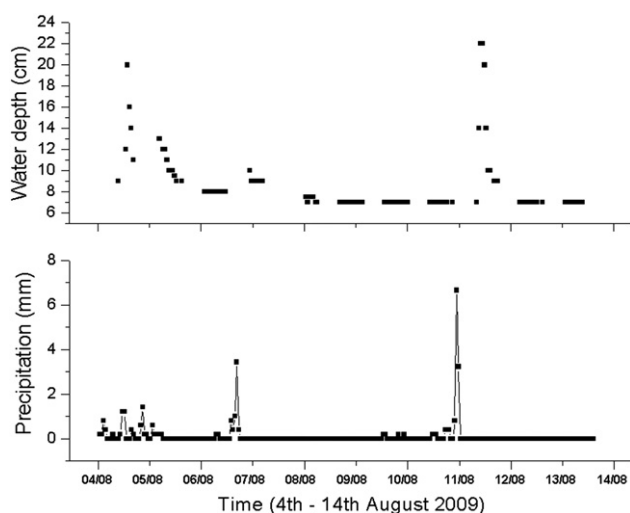
Fig. 2 – River stage and precipitation amount during 4th–14th August 2009.

hydrochloric acid. Rainfall data were obtained from Winterbourne climate station (52.455 N–1.924W, 131 m), managed by the University of Birmingham. River stage height was recorded several times a day by manual observations. The daily averaged values of all recorded parameters are presented in Table 1. The standard deviation for fluorescence intensity, on the 10th, is very high because of the pollution in the evening which increased the signal compared to previous values.

Control "grab" samples were taken every hour during the daytime, from the river, and measured for fluorescence and physical and chemical parameters (pH, conductivity and DOC). Both real-time and grab samples were measured using the same bench-top fluorometer coupled with a fibre-optic probe. The grab samples were not filtered or diluted prior to measurements. The samples were measured immediately after collection by extracting the fibre-optic probe from the real-time sample chamber, rinsing it with deionised water and inserting it into the grab sample bottle. Also, hourly grab samples were taken from the sample chamber and measured for absorbance, pH, conductivity and dissolved organic carbon (DOC). These samples were collected in order to check for possible interferences due to the experimental design such as clogging of the filter system or biofilm build-up on the fibre-optic probe or tubing.

3. Results and discussion

The experimental period was characterized by two major precipitation events and one significant pollution event. Between August 4th and 7th there were repeated short-duration rainfall events, with stage varying from 20 cm to 8 cm, within the normal range for this river. On August 12th there was a heavy short-duration rain event with precipitation quantities of 7 mm. The precipitation amount and river stage for the entire sampling period, 4th–14th August 2009, are presented in Fig. 2. The pollution event was recorded on the 7th day of measurements (August 10th). According to the

Environment Agency, an unknown quantity of diesel oil was dumped ∼1.5 km upstream. This pollution generated a visible odour and patches of oil were visible on the water surface for the first ∼24 h.

Diesel pollution appeared in the same optical space, on the EEM, as tryptophan (peaks T_1 and T_2), but slightly shifted to lower emission wavelengths (∼340 nm). Laboratory studies performed by Divya and Mishra (2008) on commercial diesel and by Li et al. (2004) on EPA refined oils showed that diesel presents fluorescence emission between 350 nm and 400 nm, but as demonstrated by Divya and Mishra (2008) the emission wavelength is highly dependent on concentration due to inner filter effects and also on the composition of the specific diesel product. Diesel fluorescence is given by the naturally fluorescent compounds polycyclic aromatic hydrocarbons (PAHs) (JiJi et al., 1999) and the higher the number of aromatic rings contained, the higher the emission wavelength. So, according to Pharr et al. (1992) and Abbas et al. (2008) two aromatic ring compounds, like naphthalene, show a fluorescence peak between 310 and 330 nm and three aromatic ring components, like phenanthrenes, between 345 nm and 355 nm. Studies made by Giamarchi et al. (2000), Baker and Curry (2004) and Selli et al. (2004), lead us to the conclusion that the fluorescence at $\lambda_{excitation}/\lambda_{emission}$ ∼225/∼340 nm (T_2) is given by the PAH, naphthalene. The second peak at $\lambda_{excitation}/\lambda_{emission}$ ∼280/∼340 nm (T_1) may correspond to phenanthrenes, as shown by Pharr et al. (1992) or, according to Giamarchi et al. (2000), the peak can relate to the presence of fluoranthene.

3.1. Comparison of real-time and control samples

For data validation and quality control, real-time measurements have been compared with measurements performed on grab samples collected directly from the river. Good correlation coefficients (Pearson's product correlation) had been calculated for the measured parameters, between control and real-time samples: peaks T_1 – 0.95, T_2 – 0.88, C – 0.79 and A – 0.59; pH – 0.52; conductivity – 0.92; DOC – 0.83. Real-time fluorescence data exhibited intensity values the same as the control samples, apart from the period with the pollution event, when the fluorescence intensity of peaks T_1 and T_2 in the control samples were higher than real-time. The control samples were integrated grab samples which therefore contained a greater proportion of the surface diesel film compared to the submerged filter. Good agreement occurred between control and real-time samples for electrical conductivity and DOC, and DOC was observed to increase at the same time that fluorescence peaks A and C intensity increased. For pH samples, control and real-time samples started to diverge after about 7 days. Sample chamber pH was 1–1.5 pH units more alkaline than the river sample by the end of the experiment, but we note that this does not lead to a divergence of fluorescence intensity data over the same time period. This fact was concluded from the similarity in fluorescence values for the control and real-time samples (as seen in Fig. 3).

3.2. Real-time fluorescence results

During the course of the continuous fluorescence measurements, 2597 fluorescence excitation–emission matrices were
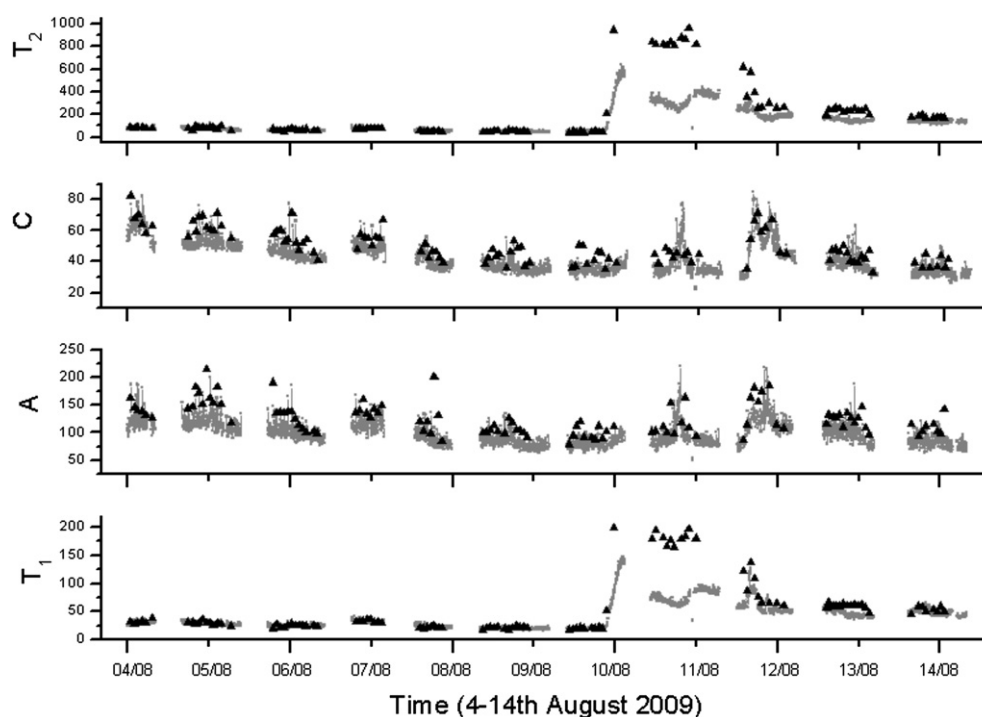
Fig. 3 – Fluorescence real-time data for peaks T_1, T_2, A and C. Grey squares – real time data; black triangles – grab samples.

collected. Only the humic substance components, peaks A and C, and tryptophan-like fluorescence maxima, T_1 and T_2 were clearly visible in the fluorescence spectra. Peak B presented very low intensity and poor discrimination from noise. Therefore the fluorescence intensity for peaks A, C, T_1 and T_2, measured using semi-automated peak-picking software, are presented in Fig. 3, together with the hourly control samples. In the period 4th–9th of August a decrease in intensity in peaks A and C can be observed. The first rain event flushed fluorescent organic matter into the river which gradually reduced until 9th of August. The same trend can be observed

to have started after the second rain event, on August 12th. The results are consistent with Carstea et al. (2009) findings which showed high influence of the precipitation on the dissolved organic matter characteristics, specifically on humic-like components. High peaks A and C fluorescence was also detected on 11th but without any rainfall; the timing coincides with upstream physical inspection of the river for the pollution source which likely stirred up river sediment and released humic-like material.

Continuous fluorescence measurements enabled the real-time detection of the diesel pollution event. The first sign of
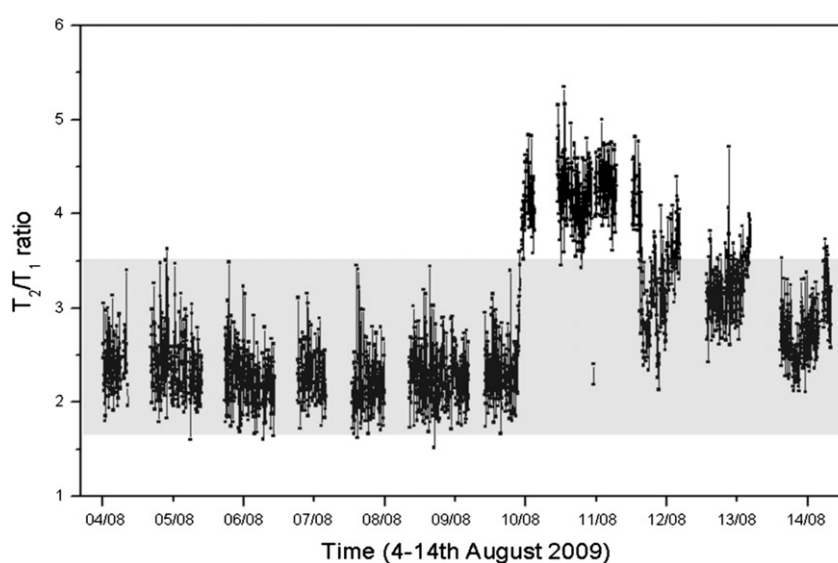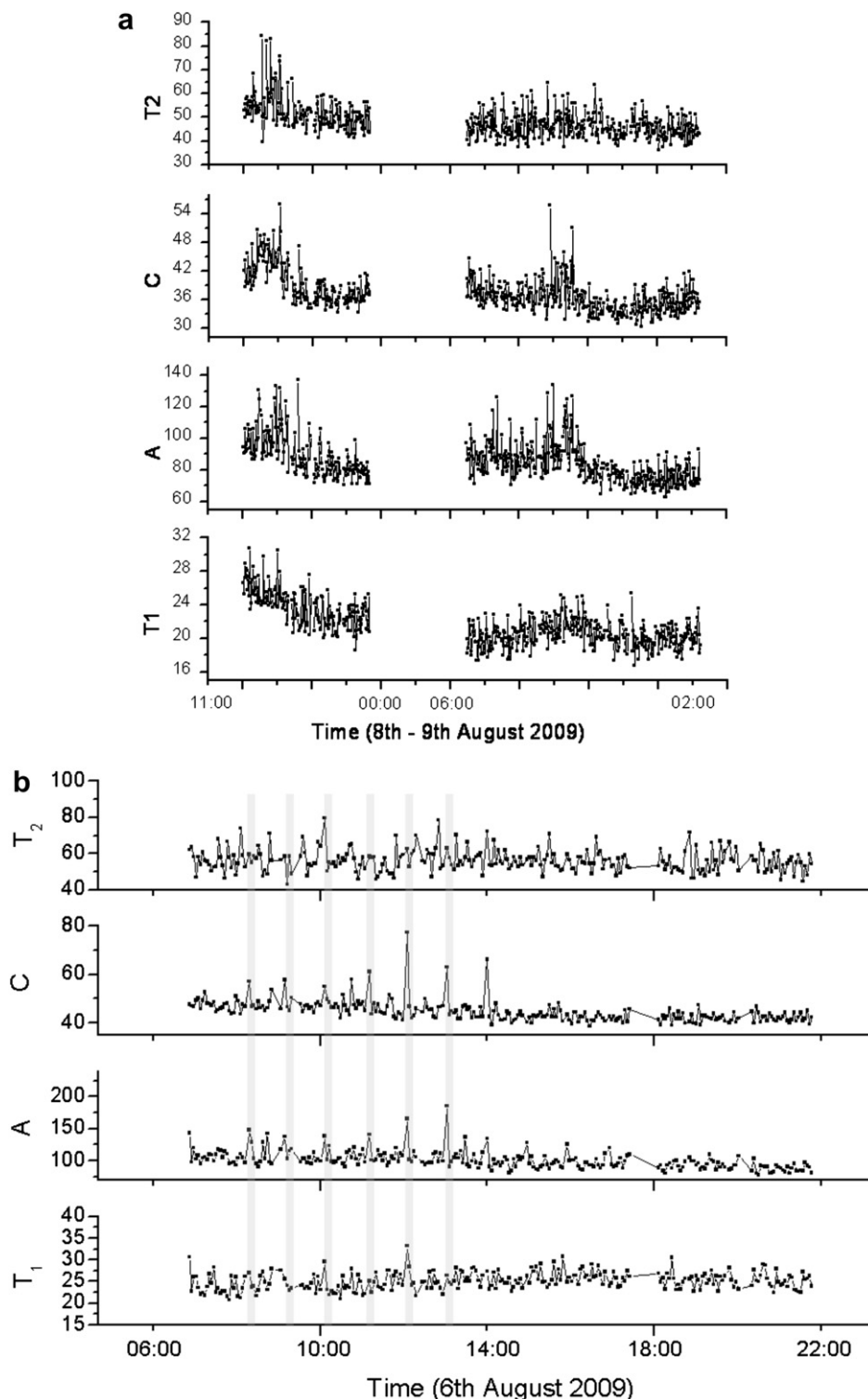

Fig. 4 – Ratio between peaks T_2/T_1 evidencing the diesel pollution.

Fig. 5 – Continuous DOM evaluation showing (a) real time data on the 8th and 9th August, showing increased variability during the daytime; and (b) approximately hourly minor pollution pulses from cross-connections observed on the 6th August.

pollution, at our sampling point, was recorded at 18:00 on the 10th August and the signal increased until 22:00. The fluorescence intensity of diesel pollution showed an initial increase followed by a decreasing intensity trend until the end of the experiment (Fig. 3). The trend was superimposed by

short-term fluctuations in intensity which correlated with the second rain event and also with the upstream river inspection, made by the Environment Agency.

The fluorescence from the diesel pollution overlapped with the tryptophan-like fluorescence excitation/emission region
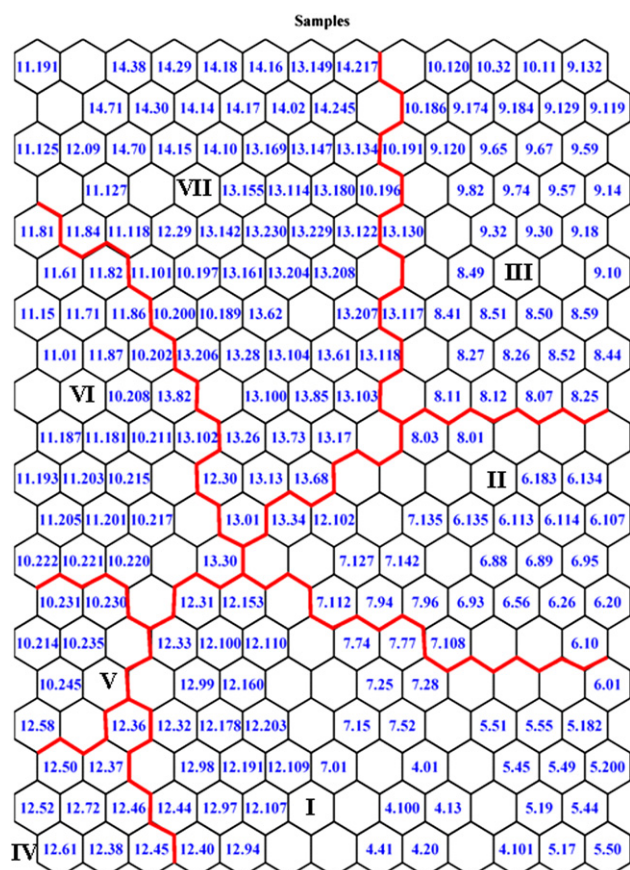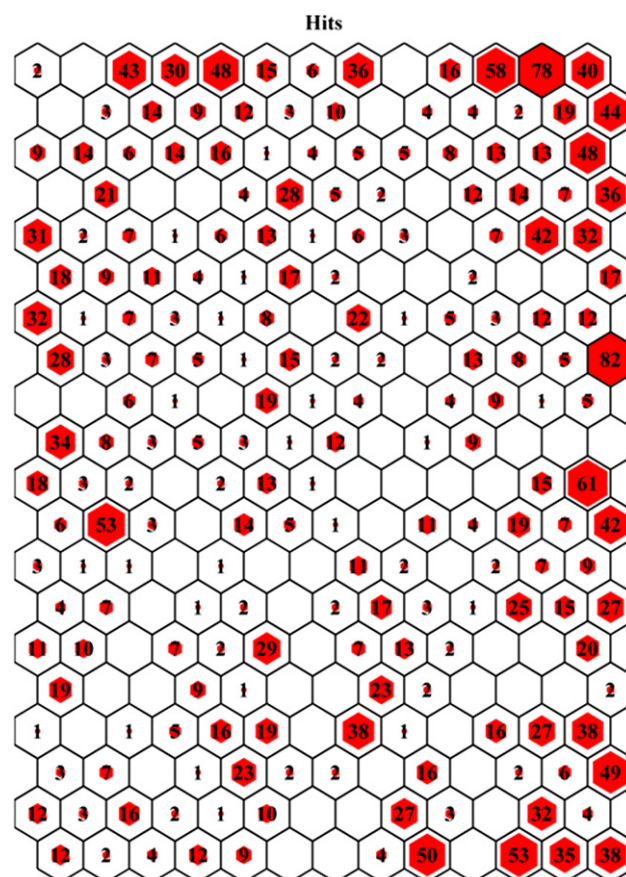
Fig. 6 – The SOM map of the continuous fluorescence data. The first number in the cell indicates the date, whereas the second number (after dot) indicates the number of the sample for a particular date (e.g. 4.41 indicates 41st sample collected on 4th August). Data clusters numbered with Roman numerals.


Fig. 7 – Hit histogram. The size of the marker and the number indicate the number of hits for particular SOM unit. Map units with the highest number of hits denote the most typical fluorescence properties in the dataset.

(indicator of sewage pollution), potentially making it difficult to separate the two pollutants. Two lines of evidence that the observed fluorescence was different from that expected from sewerage pollution are: firstly, fluorescence peak B did not increase during the pollution event and secondly, the ratio between T_2 and T_1 is different for the two contaminants. Hudson et al. (2008) observed, from ∼300 freshwater and sewerage samples, and using the same instrumentation, that the T_2/T_1 ratio is ∼3 for samples collected from sewage effluents and ∼2 for surface water samples. In our study, the values recorded before the pollution event varied between 1.7 and 3.5, whereas the fluorescence ratio during the diesel pollution was between 4 and 5. The values for T_2/T_1 ratio are presented in Fig. 4.

Closer inspection of the real time data suggest that the measurements before the diesel pollution event and outside the periods of rainfall, fluorescence intensity also show day night contrasts and hourly pulses (Fig. 5). Higher intensities and more variable fluorescence intensity can be observed during the day, and less variable data and slightly lower fluorescence intensities at nighttime. This trend is observed daily, during the study period. We speculate that these minor pollution pulses can be attributed to cross-connections with sewer systems:

upstream from our sample site cross-connected storm sewers are visible which connect to the adjacent University and hospital. These likely discharge more frequently during the working day and suggest a powerful application of real-time fluorescence monitoring in pollution monitoring. A daily trend, in the fluorescence of dissolved organic matter, was also observed by Spencer et al. (2007) and Downing et al. (2009) on measurements performed at 30 min intervals. They recorded higher fluorescence intensities during the day and lower values at nighttime, but no hourly pollution pulses were noticed. Contrary to these studies, Carstea et al. (2009) found no significant daily variability on samples collected from the Bourn Brook, but it must be noted that their fluorescence measurements were made at hourly scale and during a rainy period. Thereby, our study shows that on-line fluorescence monitoring should be performed at higher frequency scale in order to detect daily variability, short pollution pulses and the location and time of major pollution events.

3.3. Exploratory analysis of EEMs using self-organizing maps

To explore the large dataset for the key EEM properties, we undertook a novel approach from neural computing, that of
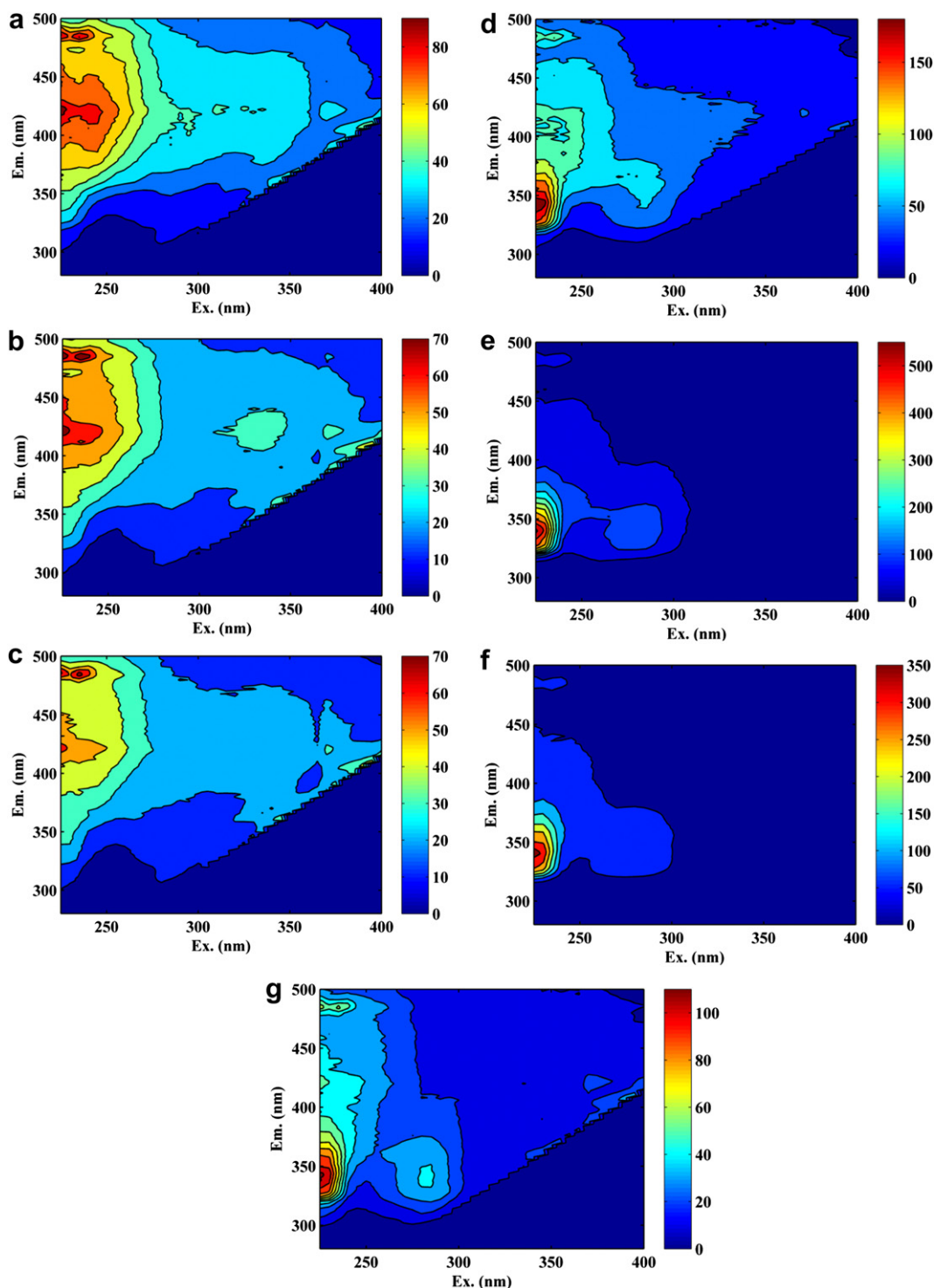
Fig. 8 − Fluorescence excitation−emission matrices for the seven SOM clusters: (a) cluster I unit 220, (b) cluster II unit 251, (c) cluster III unit 248, (d) cluster IV unit 59, (e) cluster V unit 16, (f) cluster VI unit 32, (g) cluster VII unit 8.

the SOM (Kohonen, 2001). SOM analysis was used as an alternative to peak-picking method to discriminate between fluorescence components, especially when dealing with overlapping multi-fluorophore solutions.

Distribution of fluorescence samples collected during the experiment on the SOM map is presented in Fig. 6. Based on the fluorescence features of the samples, the SOM map has

been divided into seven clusters, each with different peculiarities. The borders of the clusters are differentiated by bold lines. The SOM map shows that the samples located at the bottom of the map have higher organic matter content and fluorescence intensities compared to the samples located in the upper part. Moreover, vertically the SOM map can be divided into two time periods: the first period between 4th and

9th August in the right-hand side of the map, and the second period between 10th and 14th August in the left-hand side of the map. The first period corresponds to measurements made before the diesel pollution event whereas the second one belongs to the pollution event.

To correlate these quantitative changes in the sample distribution with the presence of particular fluorescence peaks and distinctive spectral properties, the distribution of samples on the SOM map can be portrayed with hit histograms (Fig. 7; Kohonen, 2001). For each map unit the hit characteristic is calculated on the basis of the map response to the input data. The map units with higher number of hits represent greater number of fluorescence samples. Therefore, the map units with the highest number of hits demonstrate the most typical fluorescence properties observed during the study. From Fig. 7 it can be seen that the most important map units are located at the edges of the map: unit $32-53$ hits, unit $220-53$ hits, unit $221-78$ hits, and unit $248-82$ hits. For each map unit its spectral properties can be inferred from numerical characteristics called the reference vectors. Thus the spectral properties derived from the reference vectors of those units can provide important information on the dominant fluorescence features of the dataset.

The EEMs typical for the seven map clusters generated by SOM are given in Fig. 8. In the right-hand side of the EEMs from Fig. 8a–c the predominance of the humic-like fluorescence (peak A and peak C) is an indicative of fluorescence properties before the pollution event. In contrast, Fig. 8d–g shows that tryptophan-like fluorescence dominated samples are typical for the diesel oil pollution event. As can be observed in Fig. 8e for cluster **V** fluorescence intensities for the peak T_1 are distinctively higher compared to the neighbouring clusters (2 times higher compared to cluster **VI** (Fig. 8f) and 4 times higher compared to cluster **IV** (Fig. 8d) and cluster **VII** (Fig. 8g)). The EEM of unit 220 compared to unit 248 (Fig. 8a and c) shows higher fluorescence intensities in all fluorescence regions. The fluorescence signature of the clusters II and III (Fig. 8b and c) demonstrates more degraded character of both peak A and peak C fluorophores compared to the cluster I (Fig. 8a).

SOM technique demonstrated here is a robust fluorescence analysis approach. In SOM analysis fluorescence data is projected onto two-dimensional lattice while preserving all topological and geometric properties of the original data. Therefore, SOM enables direct comparison between fluorescence samples. The main advantage over commonly used in fluorescence studies techniques for fluorescence data modelling, e.g. PCA or PARAFAC is lack of time-consuming pre-processing and validation procedures. While peak-picking approach characterizes only small portion of fluorescence EEM and in arbitrary manner selects the components of interest, SOM is an entirely data-driven approach and no assumptions have to be made regarding spectral location of fluorophores. Analysis of the entire EEMs in the SOM approach enables detection of regularities in complex mixtures of fluorophores that could be overlooked by simple analysis of selected peaks. High noise and fault tolerance of SOM makes it adequate tool for on-line analysis of fluorescence data without pre-processing, i.e. removing of the scatter. An expert system based on SOM can provide on-line quantitative and qualitative evaluation of fluorescence data and effective detection of organic pollutants, e.g. diesel pollution.

4. Conclusions

- Continuous real-time fluorescence and water quality measurements were performed on samples collected for 11 days through a tubing and filter system from the Bourn Brook River. This was the first study of on-line and in-situ monitoring of an urban river. The study proved that the system can run continuously without any cleaning for at least two weeks. Further studies are needed to test how long the system can measure before significant fouling issues arise.
- Apart from increased observations of biofilm fragments, as well as a gradual increase in sample chamber pH, no major problems were encountered and the system ran until a power supply failure occurred. The increased biofilm and pH values had no effect on the real-time fluorescence analyses, showing similar values with control samples.
- Fluorescence peak-picking method showed that humic-like fluorescence (peak C) had high intensities in the first days of the experiment, generated by the precipitation events in that period, compared to protein-like fraction. Peaks A and C also seemed to be more variable in samples measured during the day and hourly pollution pulses were recorded due to cross-connections with the sewer system.
- One diesel pollution event was detected with fluorescence spectroscopy, but it overlapped the fluorescence of the tryptophan-like component. Despite this, it formed a distinctive SOM unit and with peak-picking only the T_2/T_1 ratio distinguished diesel pollution (ratios of 4–5 for diesel pollution compared to values from 1.75 to 3.5 at other times).
- This study proved that fluorescence monitoring at high frequency, of less than 30 min, is preferred for dissolved organic matter variability assessment and for quick detection of a particular contaminant, like sewage or petroleum derived products. In case of data analysis, peak-picking method can be applied very well together with SOM analysis in order to evaluate the daily characteristics of dissolved organic matter components and discriminate between overlapping fluorophores. An expert system incorporating SOM can be applied for the on-line evaluation of fluorescence data and early warning system of failure of water quality in natural and engineered water systems.

Acknowledgements

REFERENCES

Abbas, O., Rebufa, C., Dupuya, N., Permanyer, A., Kister, J., 2008. Assessing petroleum oils biodegradation by chemometric analysis of spectroscopic data. Talanta 75, 857–871.

Ahmad, S.R., Reynolds, D.M., 1999. Monitoring of water quality using fluorescence technique: prospect of on-line process control. Water Research 33 (9), 2069–2074.

Aymerich, I.F., Piera, J., Soria-Frisch, A., Cros, L., 2009. A rapid technique for classifying phytoplankton fluorescence spectra based on self-organizing maps. Applied Spectroscopy 63 (6), 716–726.

Baker, A., 2001. Fluorescence excitation–emission matrix characterisation of some sewage impacted rivers. Environmental Science and Technology 35, 948–953.

Baker, A., Curry, M., 2004. Fluorescence of leachates from three contrasting landfills. Water Research 38 (10), 2605–2613.

Baker, A., Spencer, R.G.M., 2004. Characterisation of dissolved organic matter from source to sea using fluorescence absorption spectroscopy. Science of the Total Environment 333, 217–232.

Barbini, R., Colao, F., Fantoni, R., Ferrari, G.M., Lai, A., Palucci, A., 2003. Application of a lidar fluorosensor system to the continuous and remote monitoring of the Southern Ocean and Antarctic Ross Sea: results collected during the XIII and XV Italian oceanographic campaigns. International Journal of Remote Sensing 24 (16), 3191–3204.

Bieroza, M., Baker, A., Bridgeman, J., 2009a. Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. Science of the Total Environment 407, 1765–1774.

Bieroza, M., Baker, A., Bridgeman, J., 2009b. Exploratory analysis of excitation-emission matrix fluorescence spectra with self-organizing maps as a basis for determination of organic matter removal efficiency at water treatment works. Journal of Geophysical Research – Biogeosciences 114, G00F07.

Bugden, J.B.C., Yeung, C.W., Kepkay, P.E., Lee, K., 2008. Application of ultraviolet fluorometry and excitation–emission matrix spectroscopy (EEMS) to fingerprint oil and chemically dispersed oil in seawater. Marine Pollution Bulletin 56, 677–685.

Carstea, E.M., Baker, A., Pavelescu, G., Boomer, I., 2009. Continuous fluorescence assessment of organic matter variability on the Bournbrook River, Birmingham, UK. Hydrological Processes 23, 1937–1946.

Chen, R.F., 1999. In situ fluorescence measurements in coastal waters. Organic Geochemistry 30, 397–409.

Cheng, W.P., Chi, F.H., Yu, R.F., 2004. Evaluating the efficiency of coagulation in the removal of dissolved organic carbon from reservoir water using fluorescence and ultraviolet photometry. Environmental Monitoring and Assessment 98, 421–431.

Conmy, R.N., Coble, P.G., Del Castillo, C.E., 2004. Calibration and performance of a new in situ multi-channel fluorometer for measurement of colored dissolved organic matter in the ocean. Continental Shelf Research 24, 431–442.

Divya, O., Mishra, A.K., 2008. Understanding the concept of concentration-dependent red-shift in synchronous fluorescence spectra: Prediction of λ_{SFS}^{max} and optimization of $\Delta\lambda$ for synchronous fluorescence scan. Analytica Chimica Acta 630, 47–56.

Downing, B.D., Boss, E., Bergamaschi, B.A., Fleck, J.A., Lionberger, M.A., Ganju, N.K., Schoellhamer, D.H., Fujii, R., 2009. Quantifying fluxes and characterizing compositional changes of dissolved organic matter in aquatic systems in situ using combined acoustic and optical measurements. Limnology and Oceanography: Methods 7, 119–131.

Drozdowska, V., 2007. Seasonal and spatial variability of surface seawater fluorescence properties in the Baltic and Nordic Seas: results of lidar experiments. Oceanologia 49 (1), 59–69.

Giamarchi, P., Stephan, L., Salomon, S., Le Bihan, A., 2000. Multicomponent determination of a polyaromatic hydrocarbon mixture by direct fluorescence measurements. Journal of Fluorescence 10 (4), 393–402.

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

Hudson, N., Baker, A., Reynolds, D., 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters – a review. River Research and Applications 23, 631–649.

Hudson, N., Baker, A., Ward, D., Reynolds, D.M., Brunsdon, C., Carliell-Marquet, C., Browning, S., 2008. Can fluorescence spectrometry be used as a surrogate for the biochemical oxygen demand (BOD) test in water quality assessment? An example from South West England. Science of the Total Environment 391 (1), 149–158.

Hur, J., Hwang, S.J., Shin, J.K., 2008. Using synchronous fluorescence technique as a water quality monitoring tool for an urban river. Water, Air, and Soil Pollution 191 (1–4), 231–243.

JiJi, R.D., Cooper, G.A., Booksh, K.S., 1999. Excitation-emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. Analytica Chimica Acta 397, 61–72.

JiJi, R.D., Andersson, G.G., Booksh, K.S., 2000. Application of PARAFAC for calibration with excitation–emission matrix fluorescence spectra of three classes of environmental pollutants. Journal of Chemometrics 14, 171–185.

Kohonen, T., 2001. Self-organizng Maps, third ed. Springer, Berlin.

Li, J., Fuller, S., Cattle, J., Way, C.P., Hibbert, D.B., 2004. Matching fluorescence spectra of oil spills with spectra from suspect sources. Analytica Chimica Acta 514, 51–56.

Mopper, K., Schultz, C.A., 1993. Fluorescence as a possible tool for studying the nature and water column distribution of DOC components. Marine Chemistry 41 (1–3), 229–238.

Patra, D., Mishra, A.K., 2002. Total synchronous fluorescence scan spectra of petroleum products. Analytical and Bioanalytical Chemistry 373, 304–309.

Pharr, D.Y., McKenzie, J.K., Hickman, A.B., 1992. Fingerprinting petroleum contamination using synchronous scanning fluorescence spectroscopy. Ground Water 30 (4), 484–489.

Reynolds, D.M., Ahmad, S.R., 1997. Rapid and direct determination of wastewater BOD values using a fluorescence technique. Water Research 31 (8), 2012–2018.

Reynolds, D.M., 2002. The differentiation of biodegradable and non-biodegradable dissolved organic matter in wastewaters using fluorescence spectroscopy. Journal of Chemical Technology and Biotechnology 77 (8), 965–972.

Rhee, J.I., Lee, K.I., Kim, C.K., Yim, Y.S., Chung, S.W., Wei, J., Bellgardt, K.H., 2005. Classification of two-dimensional fluorescence spectra using self-organizing maps. Biochemical Engineering Journal 22, 135–144.

Selli, E., Zaccaria, C., Sena, F., Tomasi, G., Bidoglio, G., 2004. Application of multi-waymodels to the time-resolved fluorescence of polycyclic aromatic hydrocarbons mixtures in water. Water Research 38, 2269–2276.

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., 2007. Diurnal variability in riverine dissolved organic matter composition determined by in situ optical measurement in the San Joaquin River (California, USA). Hydrological Processes 21 (23), 3181–3189.

Vodacek, A., Hoge, F.E., Swift, R.N., Yungel, J.K., Peltzer, E.T., Blough, N.V., 1995. The use of in-situ and airborne fluorescence measurements to determine BTV absorption-coefficients and DOC concentrations in surface waters. Limnology and Oceanography 40, 411–415.

Winter, A.R., Fish, T.A.E., Playle, R.C., Smith, D.S., Curtis, P.J., 2007. Photodegradation of natural organic matter from diverse freshwater sources. Aquatic Toxicology 84, 215–222.