

Article

## Prediction of BOD, COD, and Total Nitrogen Concentrations in a Typical Urban River Using a Fluorescence Excitation-Emission Matrix with PARAFAC and UV Absorption Indices

Jin Hur \* and Jinwoo Cho

Department of Environment & Energy, Sejong University, 98 Gunja-dong, Gwangjin-ku, Seoul 143-747, Korea; E-Mail: jinwoocho@sejong.ac.kr

\* Author to whom correspondence should be addressed; E-Mail: jinhur@sejong.ac.kr;  
Tel.: +82-2-3408-3826; Fax: +82-2-3408-4320.

Received: 6 December 2011; in revised form: 6 January 2012 / Accepted: 13 January 2012 /  
Published: 16 January 2012

---

**Abstract:** The development of a real-time monitoring tool for the estimation of water quality is essential for efficient management of river pollution in urban areas. The Gap River in Korea is a typical urban river, which is affected by the effluent of a wastewater treatment plant (WWTP) and various anthropogenic activities. In this study, fluorescence excitation-emission matrices (EEM) with parallel factor analysis (PARAFAC) and UV absorption values at 220 nm and 254 nm were applied to evaluate the estimation capabilities for biochemical oxygen demand (BOD), chemical oxygen demand (COD), and total nitrogen (TN) concentrations of the river samples. Three components were successfully identified by the PARAFAC modeling from the fluorescence EEM data, in which each fluorophore group represents microbial humic-like (C1), terrestrial humic-like organic substances (C2), and protein-like organic substances (C3), and UV absorption indices (UV<sub>220</sub> and UV<sub>254</sub>), and the score values of the three PARAFAC components were selected as the estimation parameters for the nitrogen and the organic pollution of the river samples. Among the selected indices, UV<sub>220</sub>, C3 and C1 exhibited the highest correlation coefficients with BOD, COD, and TN concentrations, respectively. Multiple regression analysis using UV<sub>220</sub> and C3 demonstrated the enhancement of the prediction capability for TN.

**Keywords:** fluorescence spectroscopy; parallel factor analysis; water quality monitoring; urban river; multiple regression analysis

---

## 1. Introduction

Continuous water quality monitoring is essential for efficient management of urban rivers and for the prompt control of pollution. Due to the rapid responses of urban rivers to intensive land use and/or diverse pollution sources, the deterioration of the water quality may be accelerated, immediately posing a direct or indirect threat to human health and aquatic ecosystems [1–3]. The degree of organic pollution which occurs due to an excessive amount of organic matter, has typically been monitored by measuring BOD and COD values in rivers. A high level of BOD deteriorates river water quality by rapid decomposition of biodegradable organic matter and the subsequent depletion of dissolved oxygen, while COD traditionally represents the total organic matter. However, both concentrations are quantified by the amount of oxygen consumed for a particular chemical oxidation of organic compounds in samples. Enrichment of total nitrogen in urban rivers may result in excessive growth of algae and macrophytes, decreased biodiversity, and odor problems [3,4].

There are several limitations of the traditional water quality parameters for use in continuous monitoring. For example, at least five days are required for the completion of the BOD measurements. The presence of toxic substances may influence the biochemical oxidation, resulting in analytical errors. Potassium dichromate, a typical oxidant for a COD test, cannot completely decompose organic matter in samples, and the degree of the chemical oxidation itself may be affected by organic matter composition and the molecular structures involved [5]. Therefore, it is necessary to develop more rapid and reliable monitoring techniques to replace the traditional water quality parameter measurements. The developed techniques are also expected to serve as a pre-screening tool to select the intensive monitoring sites for extensive urban river systems.

Biosensors based on variations in currents produced from oxygen-consuming microorganisms have been suggested as a continuous monitoring tool for organic pollution. However, their short lifetime and vulnerability to environmental factors limit their applications for continuous monitoring [6]. Recently, fluorescence spectroscopy has emerged as a useful optical sensor-based monitoring technique for organic pollution [1,7,8]. In particular, EEM spectroscopy can capture various fluorescent components contained in water samples over a wide range of excitation and emission wavelengths. In EEM, protein-like fluorescence peaks have been linked to the amount of microbially produced aromatic amino acids such as tryptophan and tyrosine, while humic-like fluorescence peaks have been assigned to the presence of condensed humified organic materials [8,9]. Hudson *et al.* [10] have shown that a tryptophan-like fluorescence peak designated at excitation/emission wavelengths of 275/340 nm could be used as an indicator of the amount of biodegradable organic matters in river water samples. Despite its usefulness as a monitoring tool, however, most of the EEM studies simply relied on “peak picking” to quantify the desired fluorescent components, neglecting the potential bias related to the spectral overlaps from a complex mixture of different fluorophores.

PARAFAC, a three way-decomposition method, has been found to be very useful in identifying the independent spectra of different types of fluorophores [11]. It is advantageous to use it as a monitoring technique beyond fluorescence EEM because it can track even small variations in EEM datasets by separating several independent groups of fluorophores from the overlapped components with a high resolution. In contrast, the weakness of PARAFAC model may include the assumption of the independence among the estimated components in the model, and potential inclusion of one or more

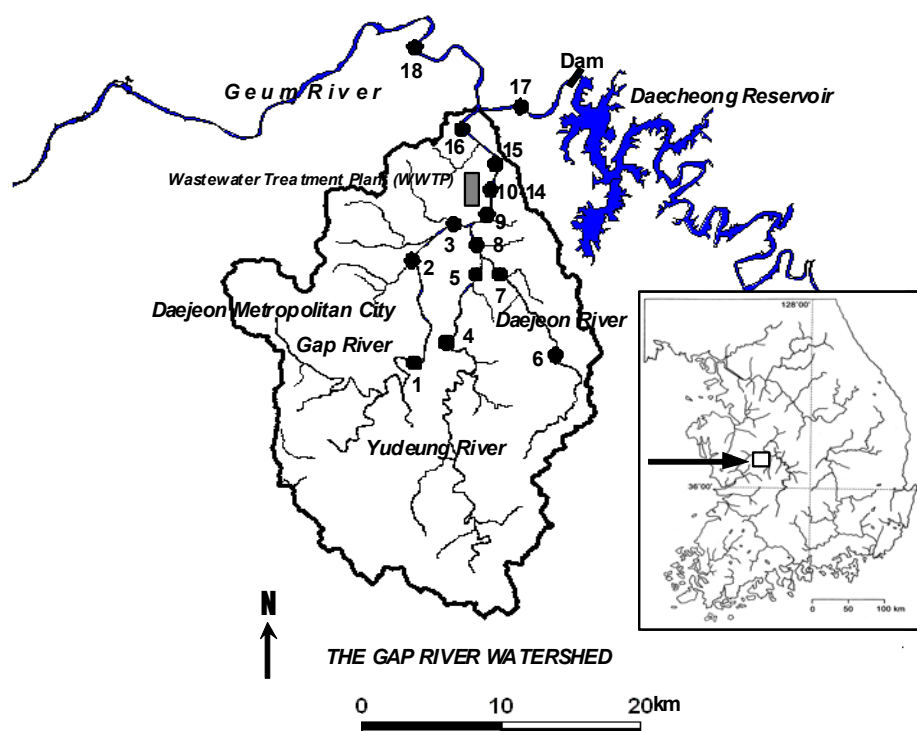
poorly estimated components, which may significantly affect the spectra and scores of all other components [12]. Most of the previous studies using PARAFAC for environmental monitoring have focused on tracking organic matter sources and/or characterizing organic matter composition [13–15]. Despite the strong implications of the close association between the fluorescence components extracted by PARAFAC and the degree of water pollution especially expressed by BOD and COD concentrations, only a few related studies have been reported [16]. Furthermore, potential applications of the PARAFAC modeling for estimating water quality parameters in typical urban rivers, which are affected by treated sewage and various anthropogenic activities, have not been fully explored. Therefore, the objectives of this study were: (1) to examine spatial variations in DOM fluorescence characteristics of a typical urban river using fluorescence EEM-PARAFAC; and (2) to estimate the degree of organic pollution and total nitrogen concentrations based on the correlations between BOD, COD, TN concentrations and the PARAFAC components.

## 2. Experimental Section

### 2.1. Study Area and Sample Collection

Water samples were collected in September and October, 2005 (09/25/2005, 10/18/2005) from eighteen locations of the Gap River watershed, a typical urban river flowing through the city of Daejeon, Korea (36°20'N, 127°26'E) with a population of 920,000. The catchment area of the watershed is 662 km<sup>2</sup> and the land use is 58% forest, 22% agriculture, and 20% urban. A municipal WWTP with a treatment capacity of 900,000 ton/day is located at the middle of the main channel of Gap River (Figure 1).

**Figure 1.** The Gap River watershed. Filled circles indicate sampling locations (adopted from Hur *et al.* [8]).



The water quality downstream of the WWTP is deteriorated by the effluent. There are two tributaries of Gap River called Yudeung River (St. 4 and 5) and Daejeon River (St. 6, 7, 8), which are located upstream of the WWTP but still affected by urban anthropogenic activities (Figure 1). Discharge from a dam reservoir (St. 17) finally joins the main channel of Gap River downstream of the WWTP. A more detailed description of the sampling locations is provided elsewhere [8]. Collected samples were kept refrigerated during transport in field before they were analyzed in a laboratory.

## 2.2. Analytical Methods

Turbidity, temperature, and pH were recorded at the sampling sites. All other analyses were made within one week after the sample collection. The collected samples were first filtered through a 0.1 mm mesh sieve to remove large sized suspended solids. The concentrations of BOD, COD, TN, and total suspended solid (TSS) were determined according to the corresponding standard methods [17]. An aliquot (50 mL) of the samples was passed through a pre-ashed GF/F filter and they were acidified with 1 M HCl to pH 3.0 for the measurements of dissolved organic carbon (DOC) and fluorescence EEM.

DOC concentrations of the samples were determined by a Shimadzu V-CPH analyzer. The relative precision of DOC measurements was <3% based on repeated measurements. Absorption measurements were performed on Varian Cary 300 Conc UV-visible spectrophotometer in a 1 cm quartz cuvette. The aliquots of the samples were diluted prior to the fluorescence measurements until UV absorbance at 254 nm was below 0.05/cm to avoid the inner-filter correction [8,18]. Fluorescence EEM of the diluted samples were generated using a luminescence spectrometer (LS-55, Perkin-Elmer) by scanning emission spectra from 300 to 550 nm at 0.5 nm increments by varying the excitation wavelength from 250 to 400 nm at 5 nm increments. Excitation and emission slits were adjusted to 10 nm and 5 nm, respectively, and the scanning speed was set at 1,200 nm/min. To limit second order Raleigh scattering, a 290 nm cutoff filter was used for all the samples. No shift of fluorescence peaks was observed by comparing the emission spectra with and without inner-filter correction for this study [8]. The fluorescence response to a blank solution (Milli-Q water) was subtracted from the EEM of each sample [19]. The measured fluorescence intensities were then standardized to a Raman peak at 395 nm emission following a suggestion by Baker [7]. Relative precisions of <2% were routinely obtained by three-times repeated fluorescence measurements of randomly chosen field samples.

## 2.3. PARAFAC Modeling

For this study, PARAFAC was applied to fully utilize the fluorescence EEM data of the samples. PARAFAC decomposes the EEM dataset into a set of trilinear terms and a residual array [18] and it estimates the underlying EEM spectra by minimizing the sum of squared residual of the trilinear model:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + \varepsilon_{ijk} \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K \quad (1)$$

where  $x_{ijk}$  is the intensity of fluorescence for the  $i$ th sample at emission wavelength  $j$  and excitation wavelength  $k$ .  $a_{if}$  is directly proportional to the concentration of the  $f$ th fluorophore in the  $i$ th sample (defined as scores).  $b_{jf}$  and  $c_{kf}$  are the estimates of the emission and excitation spectra, respectively, for

the  $f$ th fluorophore.  $F$  represents the number of components in the model and  $\varepsilon_{ijk}$  is the residual element, representing the variability not accounted for by the model [11].

PARAFAC modeling was performed using the MATLAB 7.0 (Mathworks, Natick, MA, USA) with the DOMFluor toolbox (<http://www.models.life.ku.dk>). The appropriate number of components was determined primarily based on the three diagnostic tools including residual analysis, core consistency and visual inspection of spectral shapes of each component, which are widely used by other similar studies [20,21]. The components extracted by PARAFAC represent groups of the organic fractions that exhibit similar fluorescence properties. The component scores indicate the relative concentration of the groups, not the actual concentration of a particular fluorophore. However, it is typically assumed that the scores are proportional to the concentrations of the different components [20]. In this study, the final component scores were obtained after the dilution factors of the samples were fully considered.

#### 2.4. Statistical Analyses

Regression and correlation analyses were performed using XLSTAT (Addinsoft, New York, NY, USA). Significances of the correlations in the statistics were evaluated using p-values. The total number of the data for the statistical analyses was 35.

### 3. Results and Discussion

#### 3.1. General Water Quality Parameters for Organic Matter

The concentrations of BOD and COD ranged from 0.5 mg/L to 25.4 mg/L and from 1.6 to 20.6 mg/L, respectively (Table 1).

**Table 1.** Monitoring data of turbidity, SS, BOD, COD, and TN concentrations for the Gap River watershed (September/October)<sup>a</sup>.

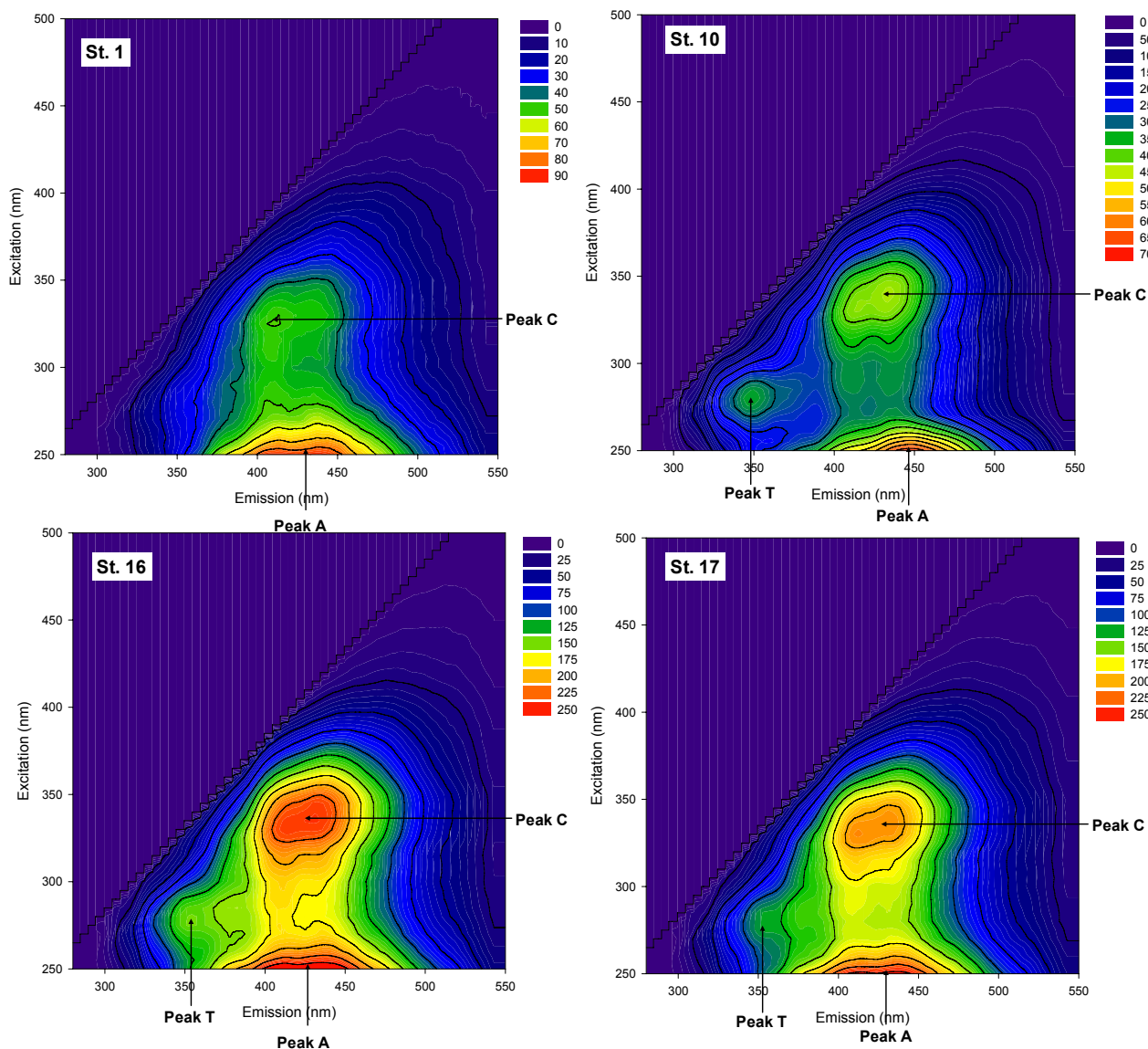
Sites	Location	Turbidity (NTU)	SS (mg/L)	BOD (mg/L)	COD (mg/L)	TN (mg/L)
St. 1	Upstream sites of the WWTP	2.5/1.4	2.3/1.6	0.7/1.1	3.2/2.6	1.9/1.9
St. 2		1.7/1.4	1.0/1.4	0.8/1.4	3.3/3.1	1.8/2.0
St. 3		5.6/3.4	6.0/3.4	1.3/1.8	4.0/4.2	2.3/2.4
St. 4		1.6/1.2	1.0/1.6	0.5/0.9	2.7/1.6	2.0/1.8
St. 5		1.2/1.5	0.5/1.6	1.0/1.2	3.6/2.3	2.0/1.9
St. 6		0.8/0.7	0.7/1.0	0.8/1.3	3.8/2.1	3.9/4.9
St. 7		1.5/1.0	1.8/1.2	1.2/1.6	4.3/3.6	4.3/4.8
St. 8		1.7/1.5	1.7/2.8	1.3/1.7	3.4/3.1	2.1/2.6
St. 9		4.3/4.1	4.8/5.4	1.4/2.4	5.7/4.9	2.3/2.1
St. 10	Near the WWTP	2.1/2.1	3.8/4.8	24.4/25.4	16.9/20.6	15.6/17.1
St. 11		3.0/ND <sup>b</sup>	4.5/ND	19.1/ND	11.7/ND	11.8/ND
St. 12		2.8/2.7	3.5/4.6	7.5/10.3	8.0/10.9	5.9/10.2
St. 13		1.6/1.7	2.5/4.2	20.2/11.7	14.2/16.4	18.4/20.0
St. 14		2.6/2.0	5.5/4.9	9.7/15.2	10.8/15.0	12.5/15.5
St. 15	Downstream sites of the WWTP	4.2/2.7	8.8/5.8	7.8/15.5	11.5/13.1	10.7/12.1
St. 16		4.1/3.7	7.0/6.4	6.1/14.2	9.0/12.1	8.0/11.8
St. 17	Discharge of a dam reservoir	2.2/2.4	2.5/3.4	1.5/1.3	5.9/5.1	1.5/1.5
St. 18	Downstream sites of the WWTP	3.3/2.8	5.3/5.6	0.6/4.0	7.0/8.1	4.4/4.6

<sup>a</sup> One sample was taken per site for each sampling event (09/25/2005, 10/18/2005); <sup>b</sup> Not determined due to the failure of the sampling.

As expected, the highest and the lowest concentrations were observed for the WWTP effluent and for the samples collected from the uppermost sites of the watershed, respectively. The concentrations of BOD exceeded the COD concentrations for the sampling sites near the WWTP, indicating that the BOD measurements may be influenced by nitrogenous BOD in sewage due to the presence of ammonia [22]. The organic matter-related water quality was substantially recovered after joining the effluent (St. 17) from the Daecheong reservoir, presenting the average BOD and COD concentrations of 2.3 mg/L and 7.6 mg/L, respectively. Total nitrogen concentrations in the watershed exhibited similar spatial variations, ranging from 1.8 mg/L to 20.0 mg/L. The levels of the three water quality parameters tend to increase downstream at the sites located upstream from the WWTP, suggesting that, aside from the WWTP effluent, the river water quality may be deteriorated by various uncontrolled pollution loads from the residential areas [23]. Additional work including the measurement of the specific pollutant loads into the main channel is required to fully justify the reasons.

### 3.2. Fluorescence EEM Characteristics

Three distinctive peaks can be identified from the fluorescence EEM of the collected samples (Figure 2). The tryptophan-like peak located at the excitation/emission wavelengths of 275 nm/340 nm (Peak T) was the most pronounced for the WWTP effluent samples (St. 10). The tryptophan-like substances may be associated with freely dissolved aromatic amino acid as well as the molecules bound with proteins and humic substances [10,24]. The peak has been used as an indicator of anthropogenic activities related to organic pollution. Many previous studies revealed a high association of the EEM peak with the presence of bioavailable and labile organic substrates and/or the product of microbial or algal activities [10,25–27]. Two types of humic-like EEM peaks were also found for all the collected samples (Figure 2). The peaks were located at the excitation/emission wavelengths of 250 nm/400–450 nm (Peak A) and 330–340 nm/350–400 nm (Peak C). The two peaks result from the presence of both carbon-carbon double bonds and aromatic carbon bonds in fulvic acid-like and/or humic-acid-like components. The relative ratio has been used as a tracer to distinguish between coastal and oceanic samples [28]. In general, higher relative ratios of the humic-like peaks to the tryptophan-like peak have been observed for river samples with low levels of organic pollution [10]. In this study, the fluorescence intensities of all the three EEM peaks exhibited an increasing trend with a higher degree of organic pollution (*i.e.*, higher BOD and COD concentrations). However, the relative peak ratios of the humic-like to the tryptophan-like fluorescence (*i.e.*, Peak A to Peak T or Peak C to Peak T) were lower for the sites near the WWTP and downstream of the WWTP compared to the upstream sites (Table 2). Our results suggest that although all the three EEM peaks tend to increase with the degree of organic pollution, Peak T may serve as a more sensitive surrogate for the point source pollution. Thus, the ratio of Peak A to Peak T or Peak C to Peak T appears to be useful as an evaluation index for the impact of treated sewage on urban rivers [8].

**Figure 2.** Fluorescence EEM spectra of the representative samples.**Table 2.** Selected spectroscopic characteristics of the river water for this study (September/October)<sup>a</sup>.

Sites	Ratio of Peak A to Peak T	Ratio of Peak C to Peak T	$A_{220}$ (l/cm)	$A_{220}-A_{254}$ (l/cm)	C1 (A.U.)	C2 (A.U.)	C3 (A.U.)
St. 1	4.39/3.30	2.41/1.88	0.44/0.45	0.41/0.43	888/899	546/475	188/221
St. 2	4.26/3.17	2.37/1.76	0.42/0.44	0.40/0.41	925/950	557/487	195/243
St. 3	4.17/3.26	2.27/1.76	0.51/0.47	0.47/0.44	1,355/1,164	760/567	286/293
St. 4	5.18/4.03	2.88/2.27	0.46/0.41	0.44/0.39	589/503	358/284	109/107
St. 5	4.31/3.39	2.30/1.95	0.46/0.41	0.44/0.39	689/623	396/337	142/145
St. 6	4.02/3.77	2.27/2.00	0.92/1.14	0.90/0.39	805/888	428/436	176/202
St. 7	3.79/3.35	2.02/1.75	0.96/1.08	0.93/1.12	964/1125	486/540	227/284
St. 8	4.04/3.08	2.22/1.63	0.46/0.54	0.44/1.06	700/821	397/414	150/210
St. 9	4.08/3.01	2.28/1.71	0.50/0.47	0.46/0.52	1,184/1,091	636/525	257/282
St. 10	1.35/2.07	1.06/1.32	1.93/1.59	1.83/2.1	5,853/6,622	3,006/4,151	3,126/3,080
St. 11	1.40/ND <sup>a</sup>	1.19/ND	1.41/ND	1.34/ND	4,036/ND	2067/ND	2,031/ND

**Table 2.** *Cont.*

St. 12	2.03/1.91	1.28/1.32	0.80/0.90	0.75/0.90	2,210/3,247	1,152/1,702	888/1343
St. 13	1.80/1.85	1.33/1.74	4.32/3.92	4.23/3.92	5,485/5,133	2,960/2,691	2,425/2,046
St. 14	2.00/1.95	1.37/1.45	2.46/2.92	2.39/2.92	4,096/3,872	2,168/2,041	1,677/1,827
St. 15	1.87/1.93	1.51/1.67	1.79/1.77	1.72/1.77	3,279/3,680	1,685/1,863	1,325/1,428
St. 16	2.26/2.18	1.55/1.78	1.54/1.84	1.49/1.84	2,486/3,301	1,212/1,644	877/1,170
St. 17	4.63/4.33	3.93/2.93	0.38/0.36	0.32/0.36	1,860/1,480	1,079/822	361/303
St. 18	3.22/2.65	2.19/1.79	0.99/0.95	0.94/0.95	1,939/2,027	1,031/1,001	511/600

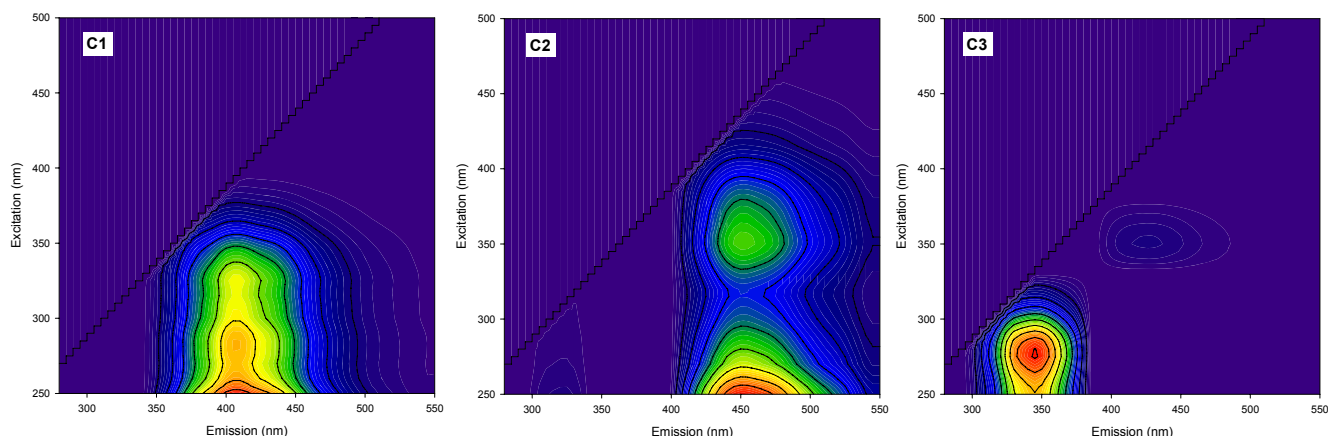
<sup>a</sup> One sample was taken per site for each sampling event (09/25/2005, 10/18/2005); <sup>b</sup> Not determined due to the failure of the sampling.

### 3.3. PARAFAC Components from EEM Data

The diagnostic tools used for this study revealed that three components are adequate for the PARAFAC model. In other words, all the fluorescence EEM data could be successfully decomposed into a three-component model by the PARAFAC analysis. Figure 3 shows each contour plots of the three PARAFAC components. A single peak, located at the excitation/emission wavelengths of 250 nm/405 nm, was observed for the contour of component 1 (C1) whereas component 2 (C2) showed two peaks at 250 nm/450 nm and at 350 nm/450 nm of the excitation/emission wavelengths. C1 and C2 may be associated with the presence of humic-like substances because the peak locations are very similar to those of Peak A and Peak C previously observed in our EEM data. However, the two fluorescent components were different from each other, not only in the number of the peaks, but also in their locations. C1 possesses a peak at a shorter emission wavelength than C2, suggesting that, despite their similar origins, the fluorophores responsible for component 1 may contain less conjugated and less condensed structures than C2 [29]. The blue-shifting of fluorescence spectra is also related to a lower degree of the humification in the DOM samples. Thus, component 1 can be interpreted as a group of less humified fluorescent substances with low molecular weights [11]. Although C1 appears to exhibit a single peak, the width of the excitation maximum (50–100 nm) indicates that it may be a mixture of fluorophores. The high ratio of Peak A to Peak C for C1 may be associated with photobleaching and/or new production of fluorophores having high excitation in Peak A region [30]. Component 3 (C3) resembles the tryptophan-like fluorescence EEM pattern, indicating the origin of the fluorophore group may be related to microbial derived amino acid and/or protein-like substances. The observed spectral characteristics of the components determined in this study generally agree with those reported in the literature based on different aquatic environments [31–33]. For particular, our PARAFAC component patterns were very consistent with a recent study by Zhang *et al.* [16], who characterized DOM in Lake Tianmuhu and its catchment basin in China using PARAFAC modeling of the fluorescence EEM data. In their study, C1 has been explained as microbial humic-like components as a result of the microbial transformation from terrestrially-derived organic matter, and C2, as a mixture of the traditional humic-like fluorescence peaks [16].



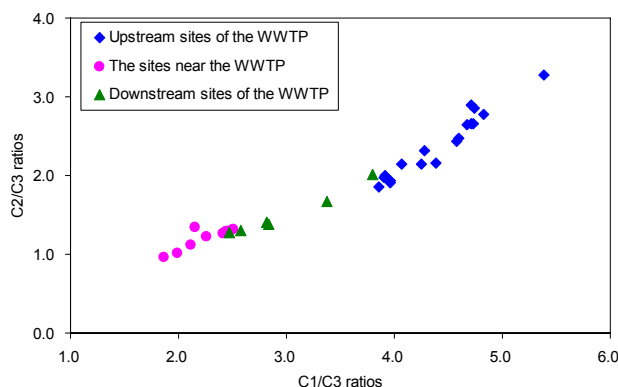
**Figure 3.** Contour plots of the three PARAFAC components decomposed from our samples.



The score values of all the components exhibited a similar trend in the variation with the concentrations of BOD and COD. The most dramatic change was observed for C3 (Table 2), suggesting that the component can be used as a good tracer for the degree of organic pollution. For the sites of the tributaries (*i.e.*, Daejeon River and Yudeung River) located upstream of the WWTP, the ratio of C1 to C2 showed an increasing trend downstream, indicating the influent rivers may continuously receive the input of less humified organic matters (*i.e.*, microbial humic-like) before they join the main channel.

A plot of the ratios of C1/C3 and C2/C3 showed spatial differences depending on the location of the sampling sites (Figure 4). For example, the sites near the WWTP exhibited relatively low ranges for the two ratios while the highest ranges were observed for the upstream sites of the WWTP. The downstream sites of the WWTP corresponded to the intermediate range. Our results suggest that C3 is highly associated with the source of the WWTP effluent. In contrast, C1 and C2 appear to reflect the characteristics of DOM in the upper sites, in which terrestrial humic-like and microbially transformed organic materials may be relatively more enriched.

**Figure 4.** A plot of the C2/C3 ratios against the C1/C3 ratios for the discrimination of the sampling sites (Site 17 is excluded).



The three components were all highly correlated each other ( $r > 0.970$ ;  $p < 0.001$ ), suggesting that the original sources of the fluorescence components are not much different from each other. As expected, the highest correlation was observed between C1 and C2 ( $r = 0.992$ ).

### 3.4. Correlations between General Organic Matter Parameters and PARAFAC Results

Correlation coefficients between some selected spectroscopic indices and water quality parameters including BOD, COD, and TN were calculated and are compared in Table 3. UV absorption values at the wavelengths of 220 nm and 254 nm, which are designated as UV<sub>220</sub> and UV<sub>254</sub>, respectively, were chosen as non-fluorescence indices because nitrate ions are known to strongly absorb UV light at a wavelength of 220 nm [15], and UV absorbance at the wavelengths between 250 nm and 280 nm has been widely used to monitor BOD and COD concentrations in river and wastewater samples [34,35]. The estimation capability of each index was evaluated using Spearman's rho as well as Pearson's r values because most of our water quality data is highly skewed and distributed in low concentration ranges.

**Table 3.** Correlation coefficients <sup>a</sup> (Pearson r values and spearman rho values) between selected spectroscopic indices and some water quality parameters (n = 35).

	<b>TN</b>	<b>BOD</b>	<b>COD</b>
UV <sub>220</sub>	0.911 <sup>b</sup>	0.706	0.770
	0.971 <sup>c</sup>	0.720	0.750
UV <sub>254</sub>	0.914	0.892	0.973
	0.747	0.813	0.954
UV <sub>220-254</sub>	0.905	0.696	0.759
	0.968	0.704	0.741
C1	0.951	0.948	0.977
	0.806	0.861	0.949
C2	0.927	0.938	0.967
	0.772	0.823	0.937
C3	0.950	0.948	0.977
	0.806	0.889	0.936

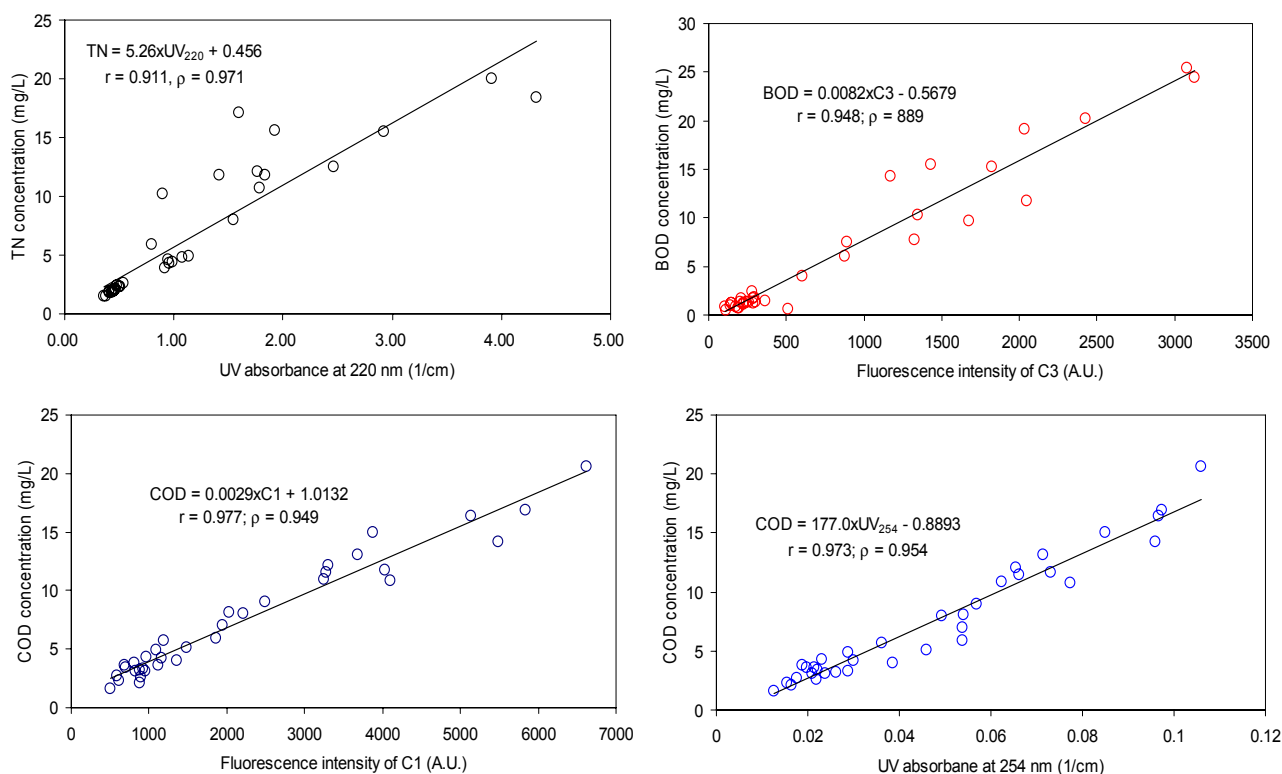
<sup>a</sup> Significant levels p values are all lower than 0.001 ( $p < 0.001$ ); <sup>b</sup> Pearson r values; <sup>c</sup> Spearman rho values.

In this study, all the selected spectroscopic indices showed significant correlations with BOD, COD and TN concentrations ( $p < 0.001$ ). As expected, TN concentrations exhibited the highest correlation with UV<sub>220</sub> based on Spearman's Rank correlation coefficient. However, such a good correlation was not very pronounced considering Pearson r values. For example, the r value based on UV<sub>220</sub> was even lower than the value calculated between UV<sub>254</sub> and TN concentrations, which did not agree with the common observation of the little UV absorption of nitrate at a wavelength of 254 nm. Our result suggests that Spearman's Rank correlation coefficients should be included in evaluating the estimation indices for a particular water quality parameter.

In this study, the difference of UV absorbance at the wavelength between 220 nm and 254 nm, UV<sub>220-254</sub>, did not enhance the estimation capability for TN concentrations (Table 3; Figure 5). The two types of the correlation coefficients (both Pearson correlation and Spearman's Rank correlation) between TN and UV<sub>220-254</sub> were both lower than those obtained based on UV<sub>220</sub>, indicating that the removal of the interference from the presence of UV-absorbing organic components did not result in the improvement of UV<sub>220</sub> estimation capability for TN. The failure of the effort may be attributed to similar trends of the variations in the organic and the nitrogen pollution, and/or similar sources of the two types of the pollution in the watershed. For example, the typical sources of nitrogen pollution in

urban areas may include fertilizer use on lawns, WWTP discharge, unattended sewage disposal, and leaks from sewer lines [36], which can also be considered as organic pollution sources. The result also suggests that nitrogen species other than nitrate ions (e.g., ammonia and organic nitrogen) may be present in an appreciable amount in the river samples.

**Figure 5.** Correlations between selected spectroscopic indices (UV absorption indices or PARAFAC components) and water quality parameters (TN, BOD, and COD).



For BOD estimation, all the PARAFAC components presented higher correlation coefficients than the absorbance-based indices, suggesting that fluorescence measurements are a superior monitoring tool for biodegradable organic matters in urban rivers. C3 exhibited the highest correlation coefficient with BOD among the PARAFAC components (Table 3; Figure 5), suggesting that microbial derived amino acid and/or protein-like substances may be good indicators for labile and biodegradable organic matters in typical urban rivers. Our results are consistent with many prior reports, in which amino acid-like or protein-like fluorescence characteristics were highly correlated with BOD concentrations in rivers [1,25,37].

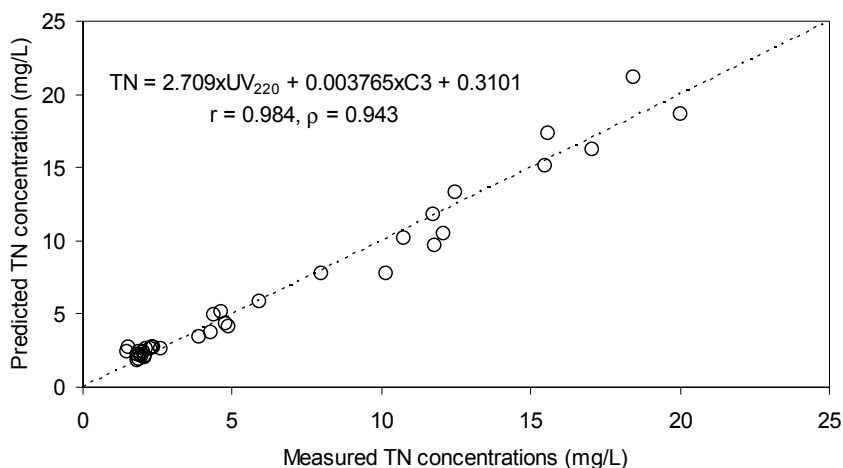
UV<sub>254</sub> showed the highest estimation capability for COD among the selected indices based on the Spearman's Rank correlation coefficients [38] while the Pearson's correlation coefficient was the highest with C1 (Table 3). C3 also presented good correlation coefficients with COD ( $r = 0.977$ ), suggesting that the protein-like component and microbial humic-like (or less humified) organic matter may constitute a dominant fraction of the total organic matter in the watershed. Our results are in contrast with other river samples not much affected polluted by sewage wastewater [16], in which COD concentrations are highly correlated with humic-like components but not with the protein-like component. Based on our individual estimation indices showing the highest correlation coefficients with TN, BOD, and COD, the regression equations are suggested as  $TN = 5.26 \times UV_{220} + 0.456$ ,

$BOD = 0.0082 \times C3 + 0.5679$ ,  $COD = 0.0029 \times C1 + 1.0132$ , and  $COD = 177.0 \times UV_{254} - 0.8893$ , respectively.

### 3.5. Improvement of the Estimation of TN Using Multiple Regression Analysis

Our previous observation of the close association between TN concentrations and the PARAFAC components (particularly C1) implies that the  $UV_{220}$  estimation capability for TN concentrations may be improved by using a multiple regression analysis based on the fluorescence indices. In this respect, a multiple regression equations based on a PARAFAC component and  $UV_{220}$  was established, and the associated correlation coefficients were calculated. Compared to the previous regression method based on a single index, the multiple regression analysis showed the enhancement of the estimation (*i.e.*, the increase in the correlation coefficients) (Figure 6).

**Figure 6.** Correlations between the measured TN concentrations and the predicted values by multiple regression method.



For example, the deviation of the data points from the regression line became much less pronounced after the multiple regression analysis was applied (Figure 6). The final equation was  $TN = 2.709 \times UV_{220} + 0.003765 \times C3 + 0.3101$ , and the correlation coefficients were 0.984 and 0.943 for Pearson's *r* value and Spearman's  $\rho$  value, respectively, exhibiting the enhancement of the estimation precision compared to the single linear correlations.

Although the fluorescence data obtained here are based on data from a standard laboratory instrument, we expect that our proposed procedure and the methodology will be useful for developing *in situ* real-time monitoring of TN, BOD, and COD concentrations in typical urban rivers. Many handheld fluorometers with various scanning function are already available, and fluorescence sensing devices are easy to make in different sizes and at a desirable level of the signal-to-noise ratio according to the purpose of the operator. More future work should be undertaken to prove successful applications of our results to *in situ* real-time monitoring techniques. For example, *in-situ* software programs to directly link the measurement data with PARAFAC modeling need to be developed. It should be noted that the regressions developed here are based on only a limited number of the samples collected during the dry season. For more successful applications, rain sampling events should be considered so that water quality and organic matter characteristics examined can be fully representative of a typical urban river.

#### 4. Conclusions

Three components were successfully identified by PARAFAC modeling from the fluorescence EEM data of the samples collected from a typical urban river, which is influenced by the effluent of a WWTP and various urban anthropogenic activities. The components represented the groups of the fluorophores representing protein-like (C3) and two humic-like organic substances (C1 and C2).  $UV_{220}$ ,  $UV_{254}$ , and the score values of the three identified PARAFAC components were chosen as the estimation indices for TN, BOD, and COD concentrations of the river samples. Among the selected indices,  $UV_{220}$ , C3 and C1 exhibited the highest correlation coefficients with TN, BOD, and COD concentrations, respectively. For TN, multiple regression analysis using the equation,  $2.709 \times UV_{220} + 0.0038 \times C3 + 0.3101$ , demonstrated the enhancement of the estimation capability. The corresponding Pearson's  $r$  values and Spearman's  $\rho$  values were 0.98 and 0.94, respectively. The selected spectroscopic indices and the associated methodology proposed here are expected to be usefully employed for the development of real-time monitoring techniques.

#### Acknowledgments

This work was supported by a National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-2011-0029028).

#### References

1. 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. *Water Res.* **2009**, *43*, 863–881.
2. Su, S.; Li, D.; Zhang, Q.; Xiao, R.; Huang, F.; Wu, J. Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China. *Water Res.* **2011**, *45*, 1781–1795.
3. Mouri, G.; Takizawa, S.; Oki, T. Spatial and temporal variation in nutrient parameters in stream water in a rural-urban catchment, Shikoku, Japan: Effects of land cover and human impact. *J. Environ. Manag.* **2011**, *92*, 1837–1848.
4. Bellos, D.; Sawidis, T.; Tsekos, I. Nutrient chemistry of River Pinios (Thessalia, Greece). *Environ. Int.* **2004**, *30*, 105–115.
5. Wang, F.; Smith, D.W.; El-Din, M.G. Aged raw landfill leachate: Membrane fractionation,  $O_3$  only and  $O_3/H_2O_2$  oxidation, and molecular size distribution analysis. *Water Res.* **2006**, *40*, 463–474.
6. Bougeois, W.; Burgess, J.E.; Stuetz, R.M. On-line monitoring of wastewater quality: A review. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 337–348.
7. Baker, A. Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers. *Environ. Sci. Technol.* **2001**, *35*, 948–953.
8. Hur, J.; Hwang, S.-J.; Shin, J.-K. Using synchronous fluorescence technique as a water quality monitoring tool for an urban river. *Water Air Soil Pollut.* **2008**, *191*, 231–243.
9. Reynolds, D.M. The differentiation of biodegradable and non-biodegradable dissolved organic matter in wastewaters using fluorescence spectroscopy. *J. Chem. Technol. Biotechnol.* **2002**, *77*, 965–972.

10. Hudson, N.; Baker, A.; Ward, D.; Reynolds, D.M.; Brunson, C.; Carliell-Marquet, C.; Browning, S. 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. *Sci. Total Environ.* **2008**, *391*, 149–158.
11. 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.
12. Murphy, K.R.; Hambly, A.; Singh, S.; Henderson, R.K.; Baker, A.; Stuetz, R.; Khan, S.J. Organic matter fluorescence in municipal water recycling schemes: Toward a unified PARAFAC model. *Environ. Sci. Technol.* **2011**, *45*, 2909–2916.
13. Stedmon, C.A.; Markager, S. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* **2005**, *50*, 686–697.
14. 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*, 40–58.
15. Bagtho, S.A.; Sharma, S.K.; Amy, G.L. Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation-emission matrices and PARAFAC. *Water Res.* **2011**, *45*, 797–809.
16. Zhang, Y.; Yin, Y.; Feng, L.; Zhu, G.; Shi, Z.; Liu, X.; Zhang, Y. Characterizing chromophoric dissolved organic matter in Lake Tianmuha and its catchment basin using excitation-emission matrix fluorescence and parallel factor analysis. *Water Res.* **2011**, *45*, 5110–5122.
17. American Public Health Association (APHA); American Water Works Association (AWWA); Water Environment Federation (WEF). *Standard Methods for the Examination of Water and Wastewater*, 20th ed.; APHA/AWWA/WEF: Washington, DC, USA, 2005.
18. Alberts, J.J.; Takács, M. Total luminescence spectra of IHSS standard and reference fulvic acids, humic acids and natural organic matter: Comparison of aquatic and terrigenous source terms. *Org. Geochem.* **2004**, *35*, 243–256.
19. Chen, W.; Westerhoff, P.; Leenheer, J.A.; Booksh, K. Fluorescence excitation-emission matrix regional integration to quantify spectra for dissolved organic matter. *Environ. Sci. Technol.* **2003**, *37*, 5701–5710.
20. Stedmon, C.A.; Bro, R. Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial. *Limnol. Oceanogr. Methods* **2008**, *6*, 572–579.
21. Li, W.-H.; Sheng, G.-P.; Liu, X.-W.; Yu, H.-Q. Characterizing the extracellular and intracellular fluorescent products of activated sludge in a sequencing batch reactor. *Water Res.* **2008**, *42*, 3173–3181.
22. Korori, H.; Ham, Y.S.; Saito, T. Influence of treated sewage effluent on organic pollution assessment in the Sakai River basin in Central Japan. *Environ. Monit. Assess.* **2009**, *151*, 243–249.
23. Cha, S.M.; Ham, Y.S.; Ki, S.J.; Lee, S.W.; Cho, K.H.; Park, Y.; Kim, J.H. Evaluation of pollutants removal efficiency to achieve successful urban river restoration. *Water Sci. Technol.* **2009**, *59*, 2101–2109.

24. Parlanti, E.; Worz, K.; Geoffroy, L.; Lamotte, M. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Org. Geochem.* **2000**, *31*, 1765–1781.
25. Baker, A.; Inverarity, R. Protein-like fluorescence intensity as a possible tool for determining river water quality. *Hydrol. Proc.* **2004**, *18*, 2927–2945.
26. Hur, J.; Park, M.H.; Schlautman, M.A. Microbial transformation of dissolved leaf litter organic matter and its effects on selected organic matter operational descriptors. *Environ. Sci. Technol.* **2009**, *43*, 2315–2321.
27. 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*, 432–535.
28. Coble, P.G. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* **1996**, *51*, 325–346.
29. Fuentes, M.; Gonzalez-Gaitano, G.; Garcia-Mina, J.M. The usefulness of UV-visible and fluorescence spectroscopies to study the chemical nature of humic substances from soils and composts. *Org. Geochem.* **2006**, *37*, 1949–1959.
30. Coble, P.G. Oceanography—Cycling coloured carbon. *Nat. Geosci.* **2008**, *1*, 575–576.
31. 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*, 7374–7379.
32. Yu, G.-H.; He, P.-J.; Shao, L.M. Novel insights into sludge dewaterability by fluorescence excitation-emission matrix combined with parallel factor analysis. *Water Res.* **2010**, *44*, 797–806.
33. Yao, X.; Zhang, Y.; Zhu, G.; Qin, B.; Feng, L.; Cai, L.; Gao, G. Resolving the variability of CDOM fluorescence to differentiate the sources and fate of DOM in lake Taihu and its tributaries. *Chemosphere* **2011**, *82*, 145–155.
34. Nataraja, M.; Qin, Y.; Seagren, E.A. Ultraviolet spectrophotometry as an index parameter for estimating the biochemical oxygen demand of domestic wastewater. *Environ. Technol.* **2006**, *27*, 789–800.
35. Fogelman, S.; Blumenstein, M.; Zhao, H. Estimation of chemical oxygen demand by ultraviolet spectroscopic profiling and artificial neural networks. *Neural Comput. Appl.* **2006**, *15*, 197–203.
36. Wakida, F.T.; Lerner, D.N. Non-agricultural sources of groundwater nitrate: A review and case study. *Water Res.* **2005**, *39*, 3–16.
37. Hur, J.; Lee, B.M.; Lee, T.H.; Park, D.-H. Estimation of biological oxygen demand and chemical oxygen demand for combined sewer systems using synchronous fluorescence spectra. *Sensors* **2010**, *10*, 2460–2471.
38. Yin, Y.; Zhang, Y.; Liu, X.; Zhu, G.; Qin, B.; Shi, Z.; Feng, L. Temporal and spatial variations of chemical oxygen demand in Lake Taihu, China, from 2005 to 2009. *Hydrobiologia* **2011**, *665*, 129–141.