



Effect of rainfall-derived inflow and infiltration on dissolved organic matter in urban sanitary sewers using chemometric and machine learning approaches

Seong-Nam Nam^{1†}, Sunghoon Lee², Jeill Oh^{2†}

¹Department of Chemical and Environmental Science, Korea Army Academy at Yeongcheon, 495 Hoguk-ro, Gogyeong-myeon, Yeongcheon-si, Gyeongsangbuk-do 38900, Republic of Korea

²Department of Civil and Environmental Engineering, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea

Received November 8, 2023 Revised July 8, 2024 Accepted July 20, 2024

ABSTRACT

This study investigated changes in dissolved organic matter (DOM) properties within urban sanitary sewers influenced by groundwater infiltration and excessive rainfall-derived inflow and infiltration (RDII). It employed optical indices and fluorescence excitation-emission matrix-parallel factor analysis (PARAFAC) coupled with self-organizing map (SOM) to compare DOM characteristics during wet weather flows (WWFs) and dry-weather flows (DWFs). Sampling sites impacted by RDII were identified based on flowrate. Optical indices and PARAFAC components (C1–C4) were used to differentiate DOM properties between DWFs and WWFs. In WWFs, E_2/E_3 and $S_{350-400}$ increased, while spectral ratio (S_R) decreased, indicating a shift towards smaller organic matters. Reduced fluorescence and humidification indices suggested the input of fresher/terrestrial organic matters. C3 and C4 exhibited significant distinctions, showing increased C3 and decreased C4 levels. The PARAFAC-SOM modeling further illustrated that water samples in the urban sewer system could be categorized based on the dominance of DOM properties. Principal component analysis revealed separation between DWF and WWF samples in principal component 1 (PC1), associated with molecular size. PC2 was linked to microbial activity in WWFs. Notably, DWF samples from the NY-11 site shifted to the positive side of the PC1 axis, while their corresponding WWF samples moved to the negative side.

Keywords: Dissolved organic matter, Fluorescence excitation-emission matrix, Parallel factor analysis, Rainfall-derived inflow and infiltration, Urban sanitary sewer system



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2025 Korean Society of Environmental Engineers

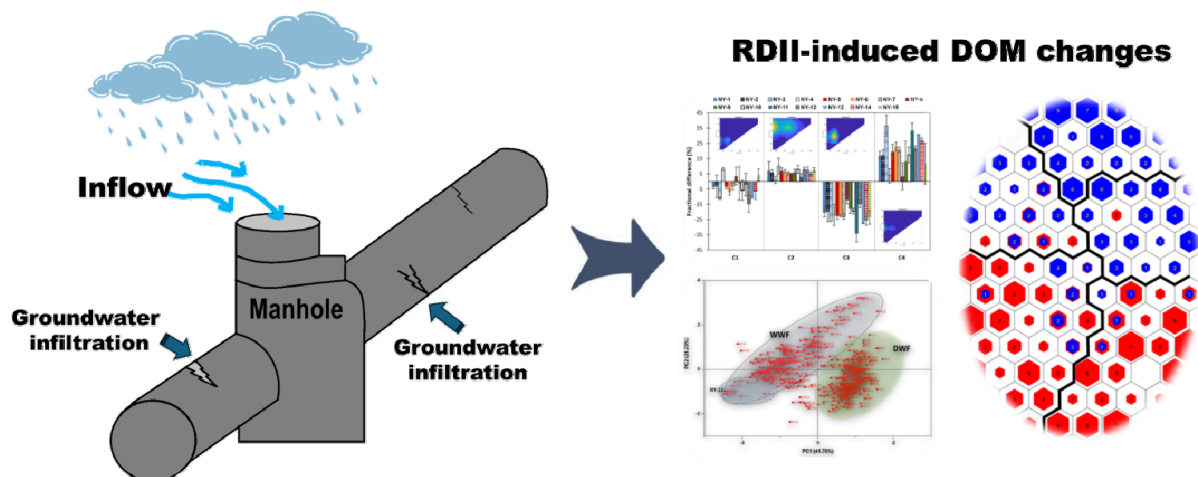
[†] Corresponding author

E-mail: namsn76@gmail.com (S.N.N.), ohjeill@cau.ac.kr (J.O.)

Tel: +82-54-330-4838 (S.N.N.), +82-2-820-5339 (J.O.)

ORCID: 0000-0002-8664-5912 (S.N.N.), 0000-0001-7321-530X (J.O.)

Graphical Abstract



1. Introduction

The flow of wastewater in a sanitary sewer system is comprised of three components, namely the base sanitary flow, groundwater infiltration, and rainfall-derived inflow and infiltration (RDII) during wet-weather flow (WWF). In most sewer systems, some inflow and infiltration (I/I) occur during dry weather flow (DWF); while small amounts of I/I in a sewer system can be tolerated, excessive I/I into capacity-constrained sewer systems may cause sanitary sewer overflows (SSOs) or bypasses [1]. RDII is stormwater entering the sanitary sewer systems in the form of inflow as well as rainfall-derived infiltration. Various sources contribute to the inflow, including direct connections (e.g., roof drains illegally connected to the sanitary sewers), surface runoff through broken manhole covers, or cross-connections between stormwater and sewer pipes [2]. Infiltration refers to groundwater that enters the sanitary sewer system through cracks or leaks in pipe sections, defective joints, and damaged manhole walls. Further, increased volumes of inflow and infiltration can dilute wastewater and change the characteristics of wastewater, which directly decreases pollutant removal efficiency of wastewater treatment plant (WWTP) operations designed for specific influent conditions. Thus, proper maintenance of wastewater collection and conveyance systems is important for effective treatment in WWTPs, prior to discharge into the environment.

Dissolved organic matter (DOM) is found almost in every type of water on earth and plays a key role in a variety of physicochemical processes and functions in aquatic environments [3-4]. For example, it serves as an energy and nutrient source for heterotrophic bacteria and some algae [5]. DOM is involved in photolytic degradation of organic pollutants as well as in pH control of aquatic systems [6]. Speciation, solubility and complexation of trace metals, and transport and fate of nanoparticles and colloids are also affected

by DOM [7-8]. The quantitative and qualitative properties of DOMs vary depending on climatologic or hydrologic conditions they are exposed to, as well as their sources [9-11]. Due to the extremely complex nature of DOM, many studies have been performed to elucidate DOM properties in quantitative and qualitative ways, using bulk characterization through fine/elemental scale methods from chromatographic analyses (size exclusion, high-resolution mass spectrometry, etc.) [12-14] to multivariate data analysis (e.g., principal component analysis (PCA), parallel factor analysis (PARAFAC)) [15-17]. Among advanced techniques, fluorescence excitation-emission matrices (EEMs) coupled with PARAFAC modeling allows us to quantitatively track the variations/dynamics in DOM as well as qualitatively differentiate its sources. The unique fluorescence properties of DOM have also been used to assess the effects of anthropogenic inputs (including surface runoff of stormwater or flow into urban sewer systems or urban streams, and studies have shown that fluorescence EEM-PARAFAC could be used to understand the influence of flows and water quality based on source-differentiated DOM [18-19].

Another promising computational method for analyzing and visualizing multi-dimensional data is the self-organizing maps (SOM) [20], which is a two-layered artificial neural network (ANN) consisting of input layer and output layer. SOM, an unsupervised machine learning technique, creates a low dimensional clustering and classification map (also called Kohonen's map) from high dimensional data [19-20].

In the present study, we coupled fluorescence EEMs obtained from multiple urban sanitary sewer points undergoing I/I with PARAFAC models to understand DOM quality under DWF and WWF conditions. The objective of this study was to assess the changes in DOM properties occurring due to RDII in urban sewer networks using EEM-PARAFAC modeling and unsupervised machine learning self-organizing map.

2. Materials and Methods

2.1. Sampling Site

Sampling sites used in the present study are in Yangju city of Kyunggi province, South Korea, and cover a catchment area of 36.80 km² with the service population of ~28,000. In this area, sewage generated is transported to the downstream WWTP with a capacity of 13,000 m³/day and treated via the DeNiPho processes. Sewage drainage in these regions was into combined sewers until 2006, when the sewers were completely replaced with separate sewer systems (with a total length of 30,038 m for sanitary sewers) to minimize I/I under dry-weather conditions. However, recent sewer inspection and flow monitoring revealed several major problems including defective pipes, false connections in pipe joints, improper manhole covers, cracks in deteriorated pipes, causing the amount of I/I to gradually increase leading to sanitary sewer overflows (SSOs). We chose a total of fifteen sampling points (Fig. S1 and Table S1), and the flow and water quality were monitored in both DWF and WWF conditions. As shown in Table S1, the studied sites have mixed land use, with a combination of residential, commercial, and industrial development.

2.2. Sewage Flowrate Measurement and Sampling

Flow rates at each point were measured once every 10 min. Data processing was performed for missing values and outliers, and the validated data were used for estimating the amount of I/I as well as average daily sewage. Fig. S2 shows the diurnal flowrate variations on the sampling dates, which were moving averaged (window size of 5) to filter noises out. At fifteen sampling points, grab samplings were conducted once every two hours for 24 h in November 2017 for dry-weather conditions and similarly in early March, 2018 for wet-weather conditions, resulting in a total of 387 samples. On the day of wet-weather flow, the rainfall intensity at the study site was 17 mm/h and the rainfall lasted for 12 h. At each sampling point, samplers collected two liters of water manually from the manhole connected to the sewers into prewashed plastic containers. Samples were kept in an ice-chest box (4°C) and transferred to the laboratory. In the laboratory, the samples were immediately filtered using a 0.45-μm membrane for including UV/Vis scanning and measurement of dissolved organic carbon (DOC) and fluorescence further analyses.

2.3. Optical Indices

To investigate the properties of DOM, several indicators from the UV/Vis scanning spectra and EEMs have been developed. Using the UV/Vis spectrum, the specific UV absorbance at 254 nm (SUVA₂₅₄) has been determined as the ratio of the absorbance at 254 nm (UV₂₅₄) to the DOC concentration and is reported to have a positive correlation with aromaticity and molecular weight of DOM [21]. Based on the UV/Vis spectrum, absorption coefficients were calculated using the following Eq. (1).

$$a(\lambda) = \frac{2.303 \text{Abs}(\lambda)}{L} \quad (1)$$

where $a(\lambda)$ represents the Napierian absorption coefficient (m⁻¹) at wavelength λ , $\text{Abs}(\lambda)$ is absorbance value at wavelength λ , and L is the path length of a cuvette (0.01 m).

The absorption ratio at 250 and 365 nm (E_2/E_3) was used as an indicator of relative molecular size of DOM, with a decreasing ratio indicating increasing molecular size, because light absorption by high molecular weight OM at longer wavelengths (i.e., at 365 nm) becomes stronger as molecular size increases [13,22]. The magnitude of spectral slopes, including $S_{275-295}$ and $S_{350-400}$, have been found to be associated with DOM modifications such as molecular weight and DOM aromaticity [22-23]. Typically, the higher values of $S_{275-295}$ and $S_{350-400}$ indicate low molecular weight (LMW) material and/or decreasing aromaticity [22,24]. Spectral slope ratio (S_R) [22] was calculated from the ratio of the spectral slopes of the absorbance between 275–295 nm ($S_{275-295}$) and 350–400 nm ($S_{350-400}$), using nonlinear regression in MATLAB 2018b (Mathworks, USA).

Fluorescence index (FI) was obtained from the ratio of fluorescence intensity at an emission wavelength of 450 nm with that at an emission wavelength of 500 nm, at an excitation wavelength of 370 nm [25]. Humidification index (HIX) indicates the degree of maturation of DOM [26]; i.e., humidification, which is positively associated with increases in the C/H ratio and the degree of aromaticity, is reflected in fluorescence intensities at longer emission wavelengths [26]. HIX was calculated as the ratio of two spectral regions at emission wavelengths 435–480 nm and 300–345 nm, at an excitation wavelength of 254 nm [26]. Biological/autochthonous index (BIX), also known as the $\beta:\alpha$ ratio, is used for estimating the degree of biological degradation of DOM. BIX was determined as the ratio of fluorescence intensity at an emission wavelength 380 nm (β peak) with the maximum intensity between the emission wavelengths 420 nm and 435 nm, at an excitation wavelength of 310 nm (α peak) [24,27].

2.4. PARAFAC and SOM Modeling

PARAFAC modeling, which is a multi-way chemometric method applicable to large-scale data organized in third- or higher-order arrays [15]. PARAFAC modeling of a three-way dataset decomposes the data signal into a set of tri-linear terms and a residual array as Eq. (2).

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (2)$$

where $i=1, \dots, I$; $j=1, \dots, J$; $k=1, \dots, K$. a_{if} (first mode) is the object score (magnitude of the fluorophore), and b_{jf} (second mode) and c_{kf} (third mode) are the excitation loading and the emission loading, respectively; e_{ijk} is the residual and contains the variation not explained by the PARAFAC model [28]. The model components have a direct chemical interpretation in a valid model. In Eq. (2), the parameter a_{if} is directly proportional to the concentration of f^{th} fluorophore in sample i ; the vectors b_{jf} and c_{kf} are scaled estimates of the emission and excitation spectra of the f^{th} fluorophore [28].

EEMs of the samples were measured using a Hitachi F-7000 fluorometer (Hitachi, Japan) with a 1-cm quartz cuvette at a constant temperature (20°C). Prior to the measurements, samples were fil-

tered through 0.45- μm membranes, and DOC in the samples was diluted down to ~ 1 mgC/L to minimize inner filter effects. EEMs were obtained by scanning over the range of 230 to 450 nm excitation wavelength at 5-nm intervals and 250 to 500 nm emission wavelength at 2-nm intervals. Thus, the dimension of each EEM consisted of 126×26 fluorescence intensity. Corrected EEMs were obtained by subtracting the EEM of ultrapure water (resistivity ≥ 18.2 M Ω -cm), followed by normalizing the EEM spectra with the area under the Raman scatter peak at the excitation wavelength of 350 nm measured on the same day as the day of sample measurement.

PARAFAC modeling was performed using the DOMFluor N-way v3.00 toolbox in MATLAB 2018b (MathWorks, USA). A total of 384 samples, with entire dataset consisting of 374 samples $\times$ 45 excitations $\times$ 126 emissions, was used for the modeling, and a two to six component PARAFAC model was used to evaluate the data to ensure that the optimum number of components was selected. All models were built using non-negativity constraints [15,28]. During PARAFAC modeling, three samples with high leverages were removed from the dataset. The optimum number of components was determined and validated using split-half analysis, core consistency and Tucker's congruence coefficient [28]. The fractional contribution (i.e., %) of each component was calculated based on the F_{max} (Raman units) of each component [15, 29].

SOM modeling was performed for combined data from PARAFAC components and optical indices in MATLAB 2022a (MathWorks, USA) using the SOM Toolbox (<http://www.cis.hut.fi/projects/somtoolbox/>) [31]. Prior to SOM analysis, the F_{max} data of the PARAFAC components has been converted to the percentage of each component. All data were normalized to values in the range of 0 to 1 with unit variance to reduce the concentration effect. According to the minimum mean quantitation error (*mqe*) and topographic error (*tge*), the optimum topology of the map

was selected to 13×8 (104 nodes) which showed the least values for both *mqe* and *tge* (Fig. S3) [19,31].

2.5. Statistical Analyses

Statistical analyses including two-sample (independent sample) t-test, and PCA were carried out using SPSS Version 26 (IBM Inc., USA). For PCA, the data were standardized by transforming variables to z-scores to minimize perversion due to different scales among variables. The rotation method was Varimax, and components with an eigenvalue >1 were used for data interpretation. The PCA results, referred to as loadings and scores, were obtained using a correlation matrix.

3. Results and Discussion

3.1. Assessment of Inflow and Infiltration

Patterns in diurnal flowrates were evaluated based on flow monitoring data collected from August 2017 to September 2018 to estimate the quantity of inflow and infiltration at each monitoring station. In dry weather, diurnal flowrates exhibited site-specific patterns, which were mirrored in land usage and service population, and were consistent with neighborhood development (e.g., high flowrates in large service areas, and vice versa) (Figs. S2 and S4). Flowrates at most of the monitoring stations near residential areas (except NY-2 and NY-11) showed evident patterns of low flows before daybreak and higher or peak flows during human activity hours (two peak flows at 6-9 AM and 7-10 PM), as shown in Fig. S2. The NY-2 site, located near City Hall and a dense commercial district, displayed relatively erratic patterns of sewage flowrate throughout the day, with increasing flows even at night. The NY-11 site, which had previously been shown to have particularly impaired sewer conditions by CCTV inspection [32], revealed no

Table 1. Estimated infiltration in DWF and RDII in WWF

Sampling site	Service area (km ²)	Designated sewage (m ³ /day)	Daily sewage in DWF (m ³ /day)			Estimated Infiltration in DWF		Estimated RDII in WWF			
			Min.	Max.	Average	Flowrate (m ³ /day)	Infiltration (%)	Flowrate (m ³ /day)	R-value	RDII (m ³)	RDII per sewer length (m ³ /m)
NY-1	0.706	210	138	283	226	31.9	14.1	470	0.011	914	0.157
NY-2	1.076	392	181	356	338	46.2	13.7	2,089	0.023	1,403	0.163
NY-3	0.575	588	276	751	473	65.5	13.8	578	0.013	562	0.27
NY-4	0.354	848	356	1,029	670	84.1	12.6	797	0.029	346	0.067
NY-5	0.929	1,437	575	1,460	1,071	128.15	12	3,809	0.017	908	0.119
NY-15	—	152	86	287	143	—	—	—	—	—	—
NY-8	1.255	6,802	1,498	6,193	4,462	118.39	2.7	3,915	0.013	936	0.103
NY-9	1.650	7,615	1,837	7,733	5,475	238.56	4.4	5,773	0.013	938	0.064
NY-11	0.414	136	173	275	225	41.28	18.3	2,370	0.031	737	0.274
NY-6	0.965	1,457	678	1,580	1,184	147.21	12.4	4,598	0.017	943	0.10
NY-7	2.041	1,849	869	2,082	1,652	180.95	11	6,340	0.024	2,346	0.126
NY-10	3.691	9,465	3,251	10,333	7,742	432.17	5.6	12,589	0.025	3,458	0.101
NY-12	4.105	9,601	3,522	11,441	8,841	465.66	5.3	14,522	0.029	4,022	0.107
NY-13	0.260	252	133	416	255	35.78	14	375	0.009	105	0.021
NY-14	0.709	560	359	812	620	83.94	13.5	975	0.007	285	0.030

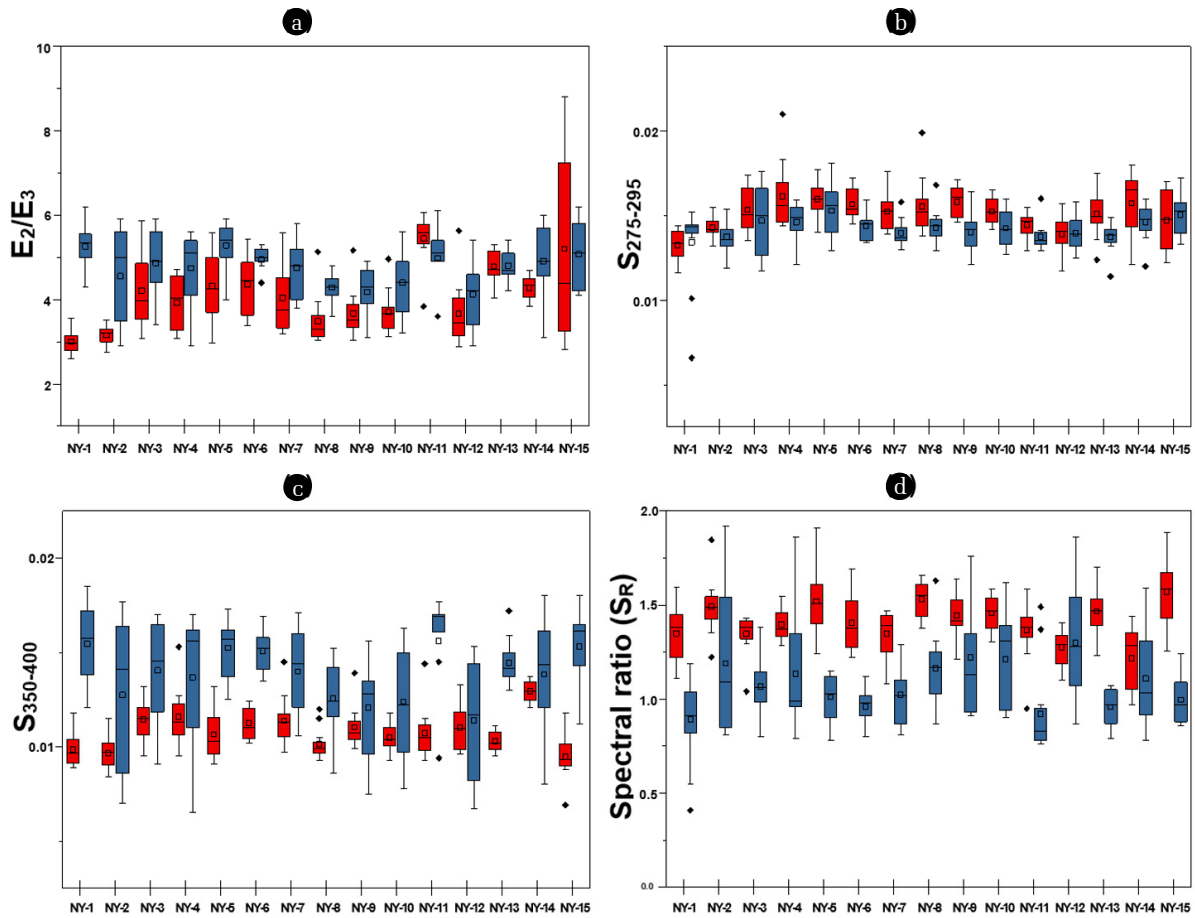


Fig. 1. Comparisons of (a) E_2/E_3 , (b)-(c) spectral slopes ($S_{275-295}$ and $S_{350-400}$), and (d) the ratio of spectral slope (S_R) in DWF and WWF conditions; red and blue boxes are DWF and WWF, respectively.

discernible trends in diurnal flowrates, and therefore ongoing groundwater penetration, depending on the water level was likely to offset flowrate changes.

Based on our previous report [32], infiltrations during the DWFs were estimated to be 2.7% to 18.3% of daily average sewage depending on the sites receiving unknown discharges including groundwater infiltration. The NY-8 and -9 sites showed the two least infiltration percentages (2.7% and 4.4%, respectively), followed by the NY-12 and -10 sites with 5.3% and 5.6% infiltration, respectively. The NY-11 site revealed the greatest infiltration (18.3% of average sewage), and the other remaining sites (NY-1~7, -13, and -14) were also found to have infiltrations of more than 11% of daily sewages. In fact, the contribution of NY-11 to the influent of the downstream WWTP ranged from negligible to ~2.5%; however, this monitoring point was suitable for distinguishing characteristic changes in organic matter due to relatively high infiltration to sewage, while the NY-8 and -9 sites served as kind of control sites for organic matter in urban sewage, because these points were less impacted by infiltration.

Based on flowrate monitoring of WWFs, the RDII volumes were calculated using the RTK method, which is the primary RDII method proposed by the United States Environmental Protection

Agency's (US EPA) stormwater management model (SWMM) [33-34]. The R parameters (the fraction of rainfall volume entering the sewer system as RDII) [1] and the RDII per unit sewer length (m^3/m) at each site were also estimated to evaluate the probable impact of RDII at the sites (Table 1 and Figs. S5a-c). As seen from the data, the greatest RDII volume was found in NY-12 followed by that in NY-10 and NY-7, and the least RDII volumes were found in NY-13, NY-14, and NY-4. In terms of the R-value and RDII volume per unit sewer length (m^3/m), NY-11 showed the greatest estimates for both calculations, indicating that sewage at the NY-11 was greatly affected by RDII throughout the entire sewer length even though the sewage volume at this site was not significant. The NY-12 showed the second greatest R-value, and a medium RDII value per sewer length (Fig. S5). Because the pipe at NY-12 was the sewer main with the largest capacity, volume increases due to RDII would be expected to significantly impact downstream WWTP performance. Thus, in terms of prioritization of sewer maintenance, NY-12 and NY-11 would need to be repaired as a matter of priority. The NY-8 and -9 sites, where the estimated infiltration % was the least in DWF conditions as mentioned above, were also affected by RDII to a relatively large extent (0.103 and 0.064 m^3 of RDII per m of sewer at NY-8 and

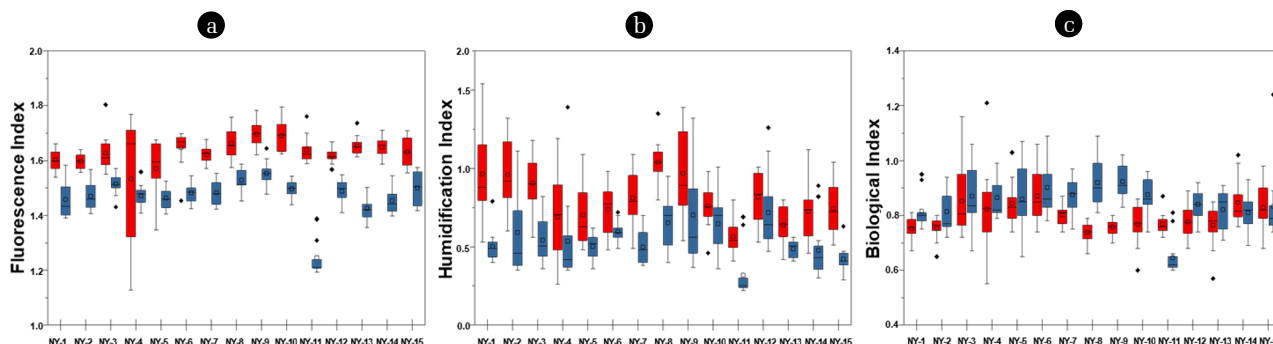


Fig. 2. Comparisons of optical indices ((a) FI, (b) HIX, and (c) BIX) extracted from fluorescence EEMs in DWF and WWF conditions; red and blue boxes are DWF and WWF, respectively.

-9, respectively). In terms of RDII per unit sewer length, NY-3 showed the second largest value ($0.27 \text{ m}^3/\text{min}$), slightly lesser than NY-11 which showed $0.274 \text{ m}^3/\text{m}$. Overall, the sewershed in the present study was found to receive substantial amounts of RDII depending on the severity of sewer conditions.

3.2. Changes of Optical Indices

In Fig. 1a, the changes of the E_2/E_3 ratio, an indicator of DOM molecular size were compared in DWFs and WWFs. In DWF conditions, the mean values of E_2/E_3 varied in the range of 2.998–5.446, and the highest E_2/E_3 was observed at the NY-11 site which showed the highest influence due to infiltration. This suggests that infiltration at this point contained higher fractions of LMW DOM, thereby causing a shift to smaller DOM molecular sizes. Meanwhile, in WWF conditions, E_2/E_3 values showed an increasing trend at most sampling sites when compared to the DWF conditions, falling in a range of 4.126–5.255. The highest value was found at NY-5 followed by NY-11 (4.985). These results imply that the RDII changed the spectrum of molecular size distribution in sewage towards LMW substances. In addition, linking to the results from PARAFAC modeling, the C3 may have a strong association with these LMW organic constituents, because the relative fraction of C3 showed a significant increase due to RDII. Both $S_{275-295}$ and $S_{350-400}$ values in DWF samples were significantly higher than those in WWF samples, with $S_{275-295}$ values ($p < 0.00$) in the range of $0.0116\text{--}0.0210 \text{ nm}^{-1}$ and $0.0066\text{--}0.0181 \text{ nm}^{-1}$ in DWF samples and WWF samples, respectively (Figs. 1b–c). The $S_{350-400}$ values ($p < 0.00$) were in the range of $0.0069\text{--}0.0153 \text{ nm}^{-1}$ and $0.0065\text{--}0.0185 \text{ nm}^{-1}$ in DWF samples and WWF samples, respectively. Overall, the mean $S_{275-295}$ and $S_{350-400}$ values were 0.0152 nm^{-1} and 0.0107 nm^{-1} in DWF samples, and 0.0142 nm^{-1} and 0.0138 nm^{-1} in WWF samples.

The spectral ratio S_R (Fig. 1d) has been also linked to shifts in DOM molecular weight, specifically, showing a negative correlation; thus, a higher S_R is an indicator of LMW [22]. In the present study, the S_R ranged between 0.949 and 1.910 (mean: 1.412), and between 0.410 and 1.920 (mean: 1.078) in DWF and WWF samples, respectively. Lower S_R in the WWF samples implies relatively increased presence of high molecular weight DOM compared to DWF samples, and this may be attributable to the substantial input of high molecular weight DOM of terrestrial origin, such as humic

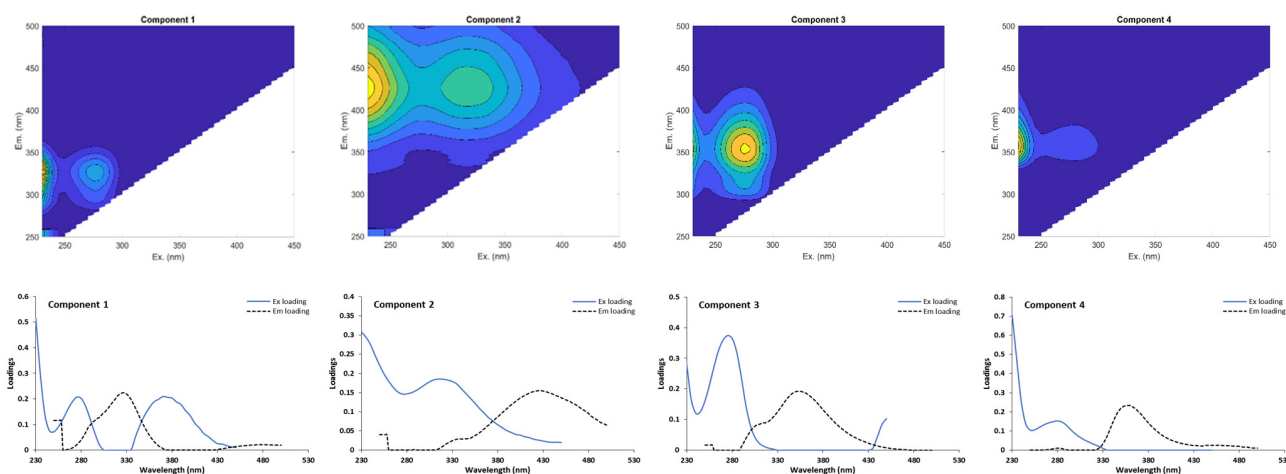
substances, due to the RDII [16].

The FI has been used as one of indicators to distinguish the autochthonous (microbially-derived) DOM from allochthonous (terrestrially sourced) DOM [25,30]. Higher FIs indicate increased autochthonous DOM while lower values mean dominance of allochthonous DOM. In previous study, for the standard reference NOMs (e.g., Suwanee River humic, fulvic acid, and NOM) derived from terrestrial sources, FIs showed values < 1 , and as for the microbially-derived DOMs such as algal OM or soluble microbial products, the values were > 1.4 [35]. High FIs exceeding 1.4 were observed in leachates in cyanobacterial intracellular organic matter (IOM) [36] and wastewater effluent organic matter (EfOM) [37–38]. FIs in natural waters were reported to be in the range of 1.2–1.8 [38–41]. In the present study, FIs varied in the range of 1.13–1.80 (mean: 1.63) and 1.19–1.64 (mean: 1.47) in DWF and WWF samples, respectively, as shown in Fig. 2a. Decreased FIs in the WWF samples are attributed to the RDII, resulting in a shift of DOM in sewage, especially towards predominance of allochthonous OM. This difference between DWF and WWF samples was also statistically significant by the two-sample t-tests ($p < 0.00$). The greatest change in the value between DWF and WWF conditions was found at the NY-11 site where the flow impacts due to groundwater infiltration and RDII were observed to be the greatest among all the sampling sites.

The HIX, an indicator of DOM humification, was compared (Fig. 2b). The mean values of HIX were found to be in the range of 0.56–1.04 and 0.32–0.72 for DWF and WWF samples, respectively. In every sampling point, the HIXs in WWF samples showed a decreasing trend compared to DWF samples, indicating that organic matter in the WWF samples was composed of relatively fresh and less mature constituents with fluorophores than that in the DWF samples. Statistical test also supported that flow conditions of DWF and WWFs contributed to the differences of HIXs ($p < 0.05$). As for the BIX values (Fig. 2c), WWF samples showed higher values than DWF samples at the majority of sampling sites (except NY-11 and NY-14), with the means of 0.74–0.87 and 0.65–0.92 for DWF and WWF samples, respectively. In correlation analysis with other parameters, its value showed moderately positive relations with FI and $S_{350-400}$ ($r > 0.6$, $p < 0.01$), especially in WWF samples (data not shown). It was reported that higher BIX values suggest that the presence of autochthonous or fresh organic

Table 2. Spectral characteristics of the four components identified using PARAFAC and comparisons to previously reported findings.

Component	λ_{ex} (nm)	λ_{em} (nm)	Description	Similar component in references
C1	<230, 275	326	Protein-like (tryptophan-like), Aromatic proteins A (or g) peak	Tryptophan-like or Amino acids in Nam <i>et al.</i> [33,42] C4 in Coble <i>et al.</i> [46] A peak in Palanti <i>et al.</i> [25]
C2	< 230, 315	426	Ubiquitous humic-like Terrestrial humic, agriculture, anthropogenic	C4 in Stedmon and Markager [14] C1 in Stedmon <i>et al.</i> [47] C2 in Cawley <i>et al.</i> [43]
C3	< 230, 275	354	Amino acids, SMP or EfOM isolates, autochthonous	C4, alumin, SMP and EfOM isolates in Nam [42] C4 in Baghoth <i>et al.</i> [44] C7 in Murphy <i>et al.</i> [45]
C4	< 230, 280	360	Protein-like substances, autochthonous	C3 in Goncalves-Araujo <i>et al.</i> [48] C4 in Cawley <i>et al.</i> [43]

**Fig. 3.** Contours of the four components identified, and their excitation and emission loadings.

matter [10] and the hydrologic conditions (e.g., the flowrates for storm periods) can be also a driver for DOM status. In this study, even though the effects on its values due to increased flowrates during the WWF event were not observed, it was assumed that increased inputs of organic matters with low molecular size and autochthonous production would be attributed to the changed BIX values.

3.3. EEM-PARAFAC Analysis of the DOM

Using PARAFAC modeling, four components were identified from the dataset consisting of DWF and WWF samples (Table 2 and Fig. 3). All the components identified have also been widely observed in other studies. One interesting observation was that the contours and spectral patterns of four components revealed multiple excitation wavelengths and one emission wavelength. Ideally, one single organic fluorophore would have one excitation and emission maxima; however, due to the complexity of natural organic matter (NOM) or DOM influenced by structures, constituents, and functional groups of humic substances and amino acids (free and bound in proteins), it is more likely to comprise a group of fluorophores [40]. The other noticeable thing in the four PARAFAC components was that all components had their ex-

citation wavelength maxima within or lower than the UV-B (280-315 nm) range, suggesting that the DOM in the sewage in the present study consisted of relatively light-resistant/refractory constituents [43].

Component 1 (C1) had excitation (Ex) maxima at 230 and 275 nm and an emission (Em) wavelength of 326 nm. The spectral location and shape of C1 resembled that of the amino acid tryptophan [44]. Component 2 (C2) showed its maxima at Ex/Em ratios of 230/426 and 315/426 nm, and its spectral pattern resembled those of humic-like substances in terrestrial sources. Components analogous to C2 have also been frequently reported by previous studies [15,45]. Component 3 (C3) and component 4 (C4) had their maxima peaks in similar regions, showing Ex/Em maxima at <230 and 275/354 nm for C3 and at <230/360 nm for C4, respectively. However, split-half analysis and component number validation clearly separated them into different components. C3 resembled amino acids from autochthonous sources [46-47] and the spectral locations of its Ex/Em were similar to those of bovine serum albumin, SMP, and EfOM isolates [44]. C4 is also thought to be originated from autochthonous fluorophores, showing spectral patterns similar to those of fulvic acids associated with microbial production and/or reworking [47].

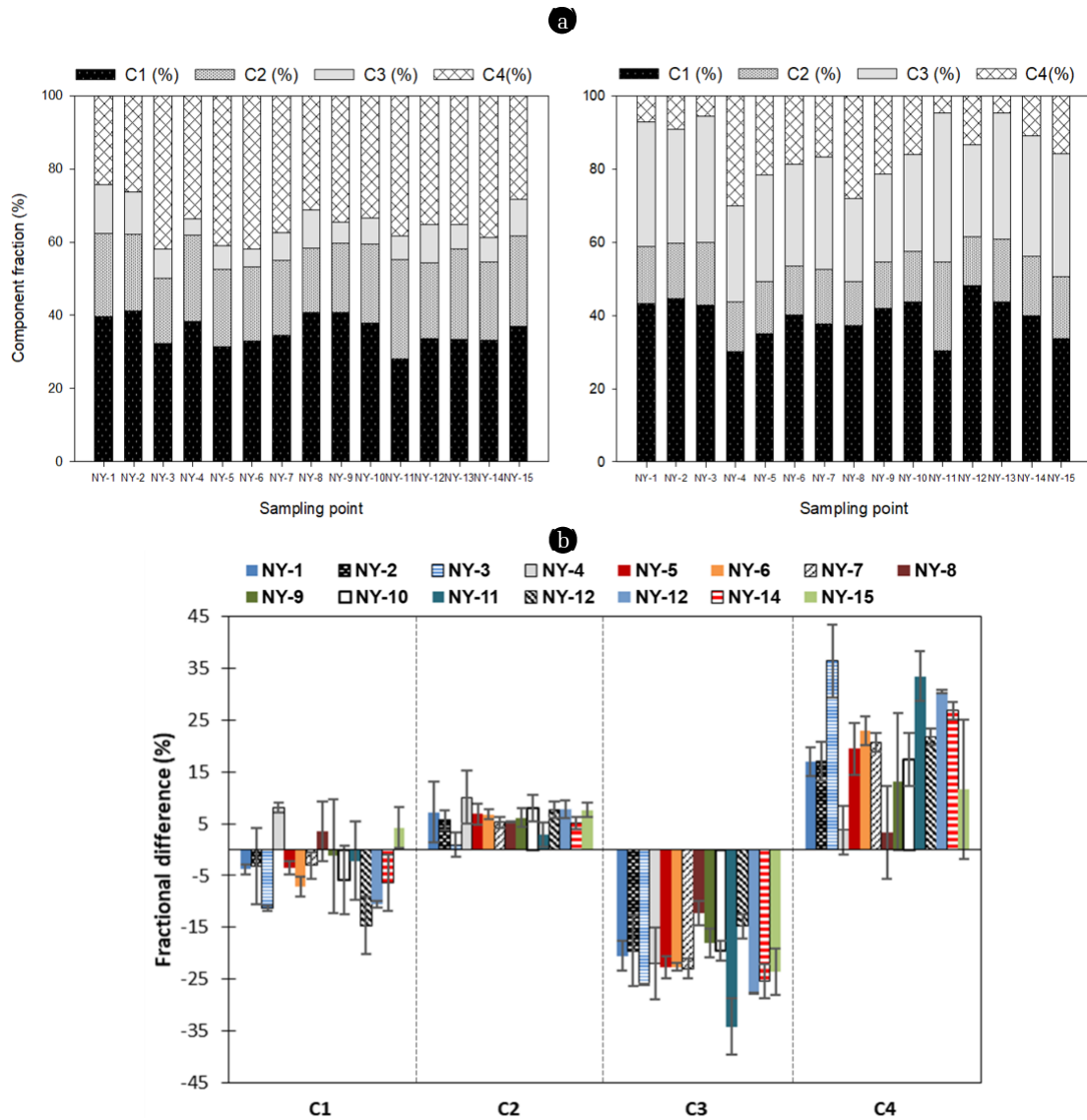


Fig. 4. (a) Fractional changes in the four components identified in dry- and wet-weather conditions; (left) DWF, (right) WWF and (b) Fractional differences between DWF and WWF. The error bar indicates standard deviation.

3.4. Component Variation by Flow Conditions

Fig. 4a shows fractional changes in the four components in dry- and wet weather conditions. Fig. 4b shows the fractional differences at each site, determined by subtracting the fractions of each component in WWFs from those in DWFs; thus, the plus values indicate decreases in fractional values in WWFs compared to those in DWFs, and the negative values indicate vice versa. Between the two flow conditions, the most noticeable changes were observed in C3 and C4, with increased fractions of the former (range: 12.2–34.2%), and decreased fractions of the latter in WWFs (range: 11.7–36.5%). This suggests that C3 is the representative fluorophore input due to RDII, and that C4 is one of the fluorophores typically found in sewage. C1, except at a few sites (NY-4, -8, and -15), showed slight increases at most of the sampling sites, and C2 showed

decreased fractions in WWFs at all sites over a range of 0.9–10.1%. Two-sample t-tests of the mean values showed that fractional changes in the four components were statistically significant in the entire sample set ($p < 0.05$) and at every site ($p < 0.05$), implying that flow conditions influenced changes in DOM.

3.5. PARAFAC-SOM Analysis

SOM analysis using the fractional compositions of PARAFAC components was performed to visually analyze the variances or similarities of sample distribution in the neuron generated by RDII event. The unified distance matrix (U-matrix), an output of SOM analysis trained on the 384-sample data (Fig. 5), depicts the distance between prototypes of fractional shares of neighboring neurons using color map units. As seen in Fig. 5a, uneven distances between

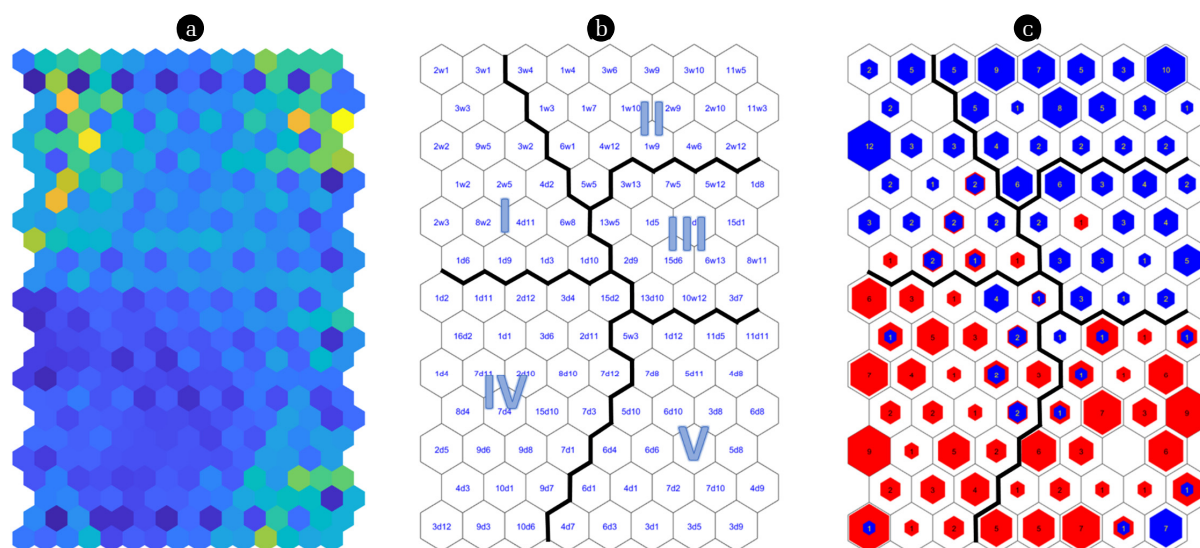


Fig. 5. U-Matrix (a), Best-matching unit (BMU) with clustering (b), hit histogram (c) for the SOM using fractional compositions of PARAFAC components for DWFs and WWFs samples.

neighbors were observed, indicating considerable dissimilarity in compositional pattern. The lower-middle portion of the map was darker and bluer than the upper portion, showing a significant difference between the upper and lower portions. Strong yellowish colors at the upper-left and upper-right corners indicate that the samples in these regions differed greatly from their neighbors. On the other hand, samples in the lower section of the U-matrix displayed relatively dark blue over a wide range, implying a high degree of similarity among samples in that region. The U-matrix can be used to assume the clustering of samples based on their gathering with similar color. However, the boundaries between clusters were not clearly displayed in the U-matrix, thus, clustering based on the Euclidean distance were identified for the entire samples into five different groups (clusters I, II, III, IV, and V) (Fig. 5b, and a dendrogram with node numbers in Fig. S6). As shown in Figs. 5b and c, by assigning the sample name to its closest output neuron (i.e., best-matching unit, BMU) on the map (13×8 topology), each sample in the original data was partitioned to the cluster. Clusters I–III were largely represented by samples of WWFs drawn in blue, with sample names beginning with the letter ‘w’, whereas clusters IV and V featured examples of DWFs beginning with the letter ‘d’, depicted in red. The dendrogram (Fig. S6), a hierarchical tree of clusters with the node number of the SOM, revealed that clusters III and IV had the greatest distance (i.e., the least similarity) between clusters whereas clusters II and III had the smallest distance (i.e., the highest similarity). The numbers marked and the size of the colored hexagons filled in the neurons (Fig. 5c) mean the number (i.e., hit) of samples falling into the winning neuron, thus, the neurons with high hits would reflect more representative features of the overall clusters. For example, in cluster I, node 3 with twelve hits showed the dominance of C1 while node 11 with nine hits presented substantially increased fraction of C4. Thus, the vertical direction from top to bottom is likely to differentiate by decreased fraction

of C1 and increased fraction of C4 from top to bottom. Meanwhile, the optical characteristics extracted from nodes 97 and 9, representing clusters III and IV, respectively, with the greatest dissimilarity, were compared. The most distinguishable discrepancies between nodes 97 and 9 were S_R and HIX, showing both optical indices in node 9 were much greater than those in node 97 (Fig. S7). High S_R is related to DOM with LMWs, thus lower S_R in samples of WWFs may indicate the inflow of DOM with high MWs. HIX is an indicator of DOM humidification. Higher HIX in node 97 than node 9 implies that DOM in WWFs is young and fresh, labile. As a result, the horizontal orientation from left to right may reflect DOM’s aging features, i.e., more aged (or old) from left to right. Meanwhile, the vertical direction of the SOM is anticipated to be dominated by shifts in major DOM components from C1 to C4, with C4 increasing and C1 decreasing in the upper to lower direction.

3.6. Principal Component Analysis

To identify factors responsible for changes of DOM quality in sewage due to excessive groundwater infiltration and RDII, PCA was applied using $\%F_{\max}$ of four components (fraction of each component in sample) along with six DOM indices ($S_{275-290}$, $S_{350-400}$, S_R , FI, HIX, and BIX). PCA was conducted on a correlation matrix of the z-score variables (i.e., mean=0, variance=1), and three principal components (PCs) (76.67% of the explaining variance of the original data) were extracted based on the criteria above an eigenvalue of 1. However, for simplicity of interpretation, only the first two PCs (PC1 and PC2) were used, accounting for 62.93% of the variance. PCA loading and score plots presented in Figs. 6a and b show that the samples were separated due to the combined effects of PC1 and PC2, which explained 41.70% and 21.23% of the entire dataset, respectively. As shown in Fig. 6a, S_R , FI, HIX and $\%C4$ showed relatively high/positive loadings in PC1 with weak loadings of $S_{275-295}$, $\%C1$, and $\%C2$, while $\%C3$ and

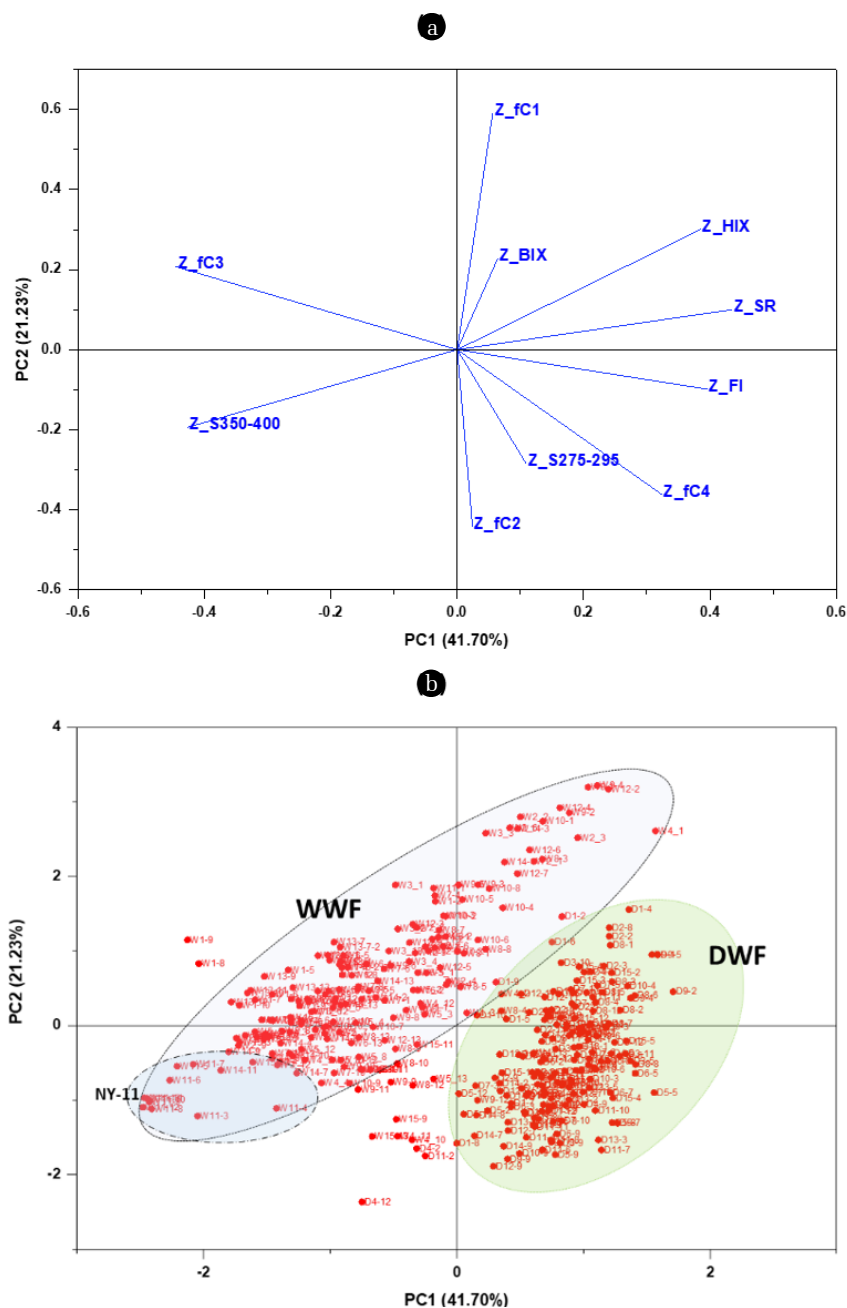


Fig. 6. PCA results. (a) Loadings of DOM indices and (b) scores of DWF and WWF samples in two principal components (PC1 and PC2).

S₃₅₀₋₄₀₀ had negative loadings in PC1. S_R and S₂₇₅₋₂₉₅ were found to be negatively related to DOM molecular size. FIs positively reflected microbial-derived organic matter. Thus, PC1 is likely to be associated with LMW-OM regardless of source. On the other hand, PC2 showed the highest positive loading in %C1, followed by HIX, BIX, %C3, and S_R. %C2, %C4, S₂₇₅₋₂₉₅, and S₃₅₀₋₄₀₀ had negative loadings in PC2. C2 represented humic-like substances of terrestrial origin, and typically presented high aromaticity. C1 and BIX were the parameters showing biological association. Therefore, PC2 may be positively associated with pollutant loading

with microbial activity. Fig. 6b shows the score scatter plot representing all the DWF and WWF samples. Samples taken in DWFs were clearly separated from those in WWFs by PC1, with DWFs showing high PC1 scores. This was consistent with high loadings of FIs, %C4, and DOM in DWFs, and although influenced by GWI, there were dominated by protein-like OM. WWFs samples showed scattered scores depending on RDII gradients. In particular, WWF samples at NY-11, showing negative loadings and scores in both PC1 and PC2, were grouped at the left-bottom, which was apparently shifted from the right-bottom. This indicates that

RDII changed the overall DOM properties in sewage at the NY-11 site towards DOM absorbing light of lower energy (i.e., longer wavelength).

4. Conclusions

In the present study, changes in DOM properties due to RDII were assessed using several optical indices, fluorescence EEMs coupled with PARAFAC and SOM modeling. The key findings are summarized as follows:

- 1) Analysis of diurnal flow rates under both DWF and WWF conditions revealed a distinctive flow pattern at the study site (NY-11), indicating significant influence from groundwater infiltration and RDII during both flow conditions.
- 2) Indices derived from UV scanning of samples, including E_2/E_3 , $S_{350-400}$, and S_R , demonstrated that RDII induces a shift in DOM constituents in sewage towards an increased fraction of LMW substances. This shift is evidenced by elevated E_2/E_3 and $S_{350-400}$ values and a decrease in S_R .
- 3) Under WWF conditions, FI and HIX values decreased, while changes in BIX were site-dependent.
- 4) The PARAFAC analysis identified four characteristic fluorophores within DOMs. Notably, C3 was identified as a potential representative influenced by RDII, whereas C4 originated from sewage. The PARAFAC-SOM modeling further demonstrated that water samples in the urban sewer system could be categorized based on the dominance of specific DOM properties.
- 5) The PCA results, utilizing both PARAFAC components and optical indices, revealed distinct separation of samples into DWF and WWF conditions. This observation underscores significant alterations in DOM properties induced by RDII. PCA showed that PC1 was associated with molecular size, while PC2 was linked to microbial activity in WWF samples.

The study's findings indicate that the intrusion of external waters (infiltration and inflow) into the sewer networks can have a negative impact on both the quantity and quality of wastewater. Increases in sewage volume induced by RDII have been connected to poor performance of receiving wastewater treatment plants (WWTPs) as well as reductions in sewers' effective hydraulic capacity. Deteriorated WWTP performance, which also leads to increased operation costs for utilities, can be attributed to either dilution of raw sewage or distinctive changes in organic constituents that deviate from the WWTP's initial design parameters.

These findings enhance our understanding of how RDII affects DOM characteristics in urban sanitary sewers, providing valuable insights for water quality management and infrastructure planning. The integration of advanced chemometric and machine learning approaches, such as PARAFAC and SOM modeling, proved to be effective in characterizing and distinguishing DOM properties under varying flow conditions.

Acknowledgements

This work was supported by the Chung-Ang University research grant in 2020, and this research was also supported by the Chung-Ang University Graduate Research Scholarship (2018).

Author Contributions

S.-N.N. (Assistant Professor) conceptualized the methodology, performed the data analysis on PARAFAC modeling, and wrote the manuscript. S.L. (M.S. student) performed the experiments and analyzed the samples. J.O. (Professor) acquired the funding and finalized the submitted version.

Conflict-of-Interest Statement

The authors declare that they have no conflict of interest.

References

1. US EPA. Review of sewer design criteria and RDII prediction methods. Available from: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRML&dirEntryId=188284.
2. Muleta MK, Boulos P. Analysis and calibration of RDII and design of sewer collection systems. World Environmental and Water Resources Congress 2008; 12 May 2008; Honolulu, Hawaii; 2008. p.1-10.
3. McLaughlin C, Kaplan LA. Biological lability of dissolved organic carbon in stream water and contributing terrestrial sources. *Freshw. Sci.* 2013;32:1219-1230. <https://doi.org/10.1899/12-202.1>.
4. Buffam I, Galloway JN, Blum LK, McGlathery KJ. A storm-flow/baseflow comparison of dissolved organic matter concentrations and bioavailability in an Appalachian stream. *Biogeochemistry.* 2001;53:269-306. <https://doi.org/10.1023/A:1010643432253>.
5. Wiegner TN, Seitzinger SP. Photochemical and microbial degradation of external dissolved organic matter inputs to rivers. *Aquat. Microb. Ecol.* 2001;24:27-40. <https://doi.org/10.3354/ame024027>.
6. Wenk J, Graf C, Aeschbacher M, Sander M, Canonica S. Effect of solution pH on the dual role of dissolved organic matter in sensitized pollutant photooxidation. *Environ. Sci. Technol.* 2021;55(22):15110-15122. <https://doi.org/10.1021/acs.est.1c03301>.
7. Yamashita Y, Jaffe R. Characterizing the interaction between trace metals and dissolved organic matter using excitation-emission matrix and parallel factor analysis. *Environ. Sci. Technol.* 2008;42:7374-7379. <https://doi.org/10.1021/es801357h>.
8. Aiken GR, Hsu-Kim H, Ryan JN. Influence of dissolved organic matter on the environmental fate of metals, nanoparticles, and colloids. *Environ. Sci. Technol.* 2011;45:3196-3201. <https://doi.org/10.1021/es103992s>.
9. Ward N, Richey J, Keil R. Temporal variation in river nutrient and dissolved lignin phenol concentrations and the impact

- of storm events on nutrient loading to Hood Canal, Washington, USA. *Biogeochemistry*. 2012;111:629-645. <https://doi.org/10.1007/s10533-012-9700-9>.
10. Guarch-Ribot A, Butturini A. Hydrological conditions regulate dissolved organic matter quality in an intermittent headwater stream. From drought to storm analysis. *Sci. Total Environ.* 2016;571:1358-1369. <https://doi.org/10.1016/j.scitotenv.2016.07.060>.
11. Inamdar S, Singh S, Dutta S *et al.* Fluorescence characteristics and sources of dissolved organic matter for stream water during storm events in a forested mid-Atlantic watershed. *J. Geophys. Res. Biogeosci.* 2011;116(G3). <https://doi.org/10.1029/2011JG001735>.
12. Petras D, Koester I, Da Silva R *et al.* High-resolution liquid chromatography tandem mass spectrometry enables large scale molecular characterization of dissolved organic matter. *Front. Mar. Sci.* 2017;4:1-14. <https://doi.org/10.3389/fmars.2017.00405>.
13. Peuravuori J, Pihlaja K. Molecular size distribution and spectroscopic properties of aquatic humic substances. *Anal. Chim. Acta.* 1997;337:133-149. [https://doi.org/10.1016/S0003-2670\(96\)00412-6](https://doi.org/10.1016/S0003-2670(96)00412-6).
14. Lee S, Park J. Comparison of molecular characteristics between commercialized and regional natural organic matters. *Environ. Eng. Res.* 2024;29:230190. <https://doi.org/10.4491/eer.2023.190>.
15. Stedmon CA, 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. <https://doi.org/10.4319/lo.2005.50.2.0686>.
16. Chen H, Liao ZL, Gu XY, Xie JQ, Li HZ, Zhang J. Anthropogenic influences of paved runoff and sanitary sewage on the dissolved organic matter quality of wet weather overflows: an excitation-emission matrix parallel factor analysis assessment. *Environ. Sci. Technol.* 2017;51:1157-1167. <https://doi.org/10.1021/acs.est.6b03727>.
17. Nguyen T, Nam SN. Sunlight-driven photocatalysis of dissolved organic matter: Tracking by excitation emission matrix-parallel factor analysis and optimization using response surface methodology. *Environ. Eng. Res.* 2021;26:200201. <http://dx.doi.org/10.4491/eer.2020.201>.
18. Shelton JM, Kim L, Fang J, Ray C, Yan T. Assessing the severity of rainfall-derived infiltration and inflow and sewer deterioration based on the flux stability of sewage markers. *Environ. Sci. Technol.* 2011;45:8683-8690. <https://doi.org/10.1021/es2019115>.
19. Zhang Y, Liang X, Wang Z, Xu L. A novel approach combining self-organizing map and parallel factor analysis for monitoring water quality of watersheds under non-point source pollution. *Sci. Rep.* 2015;5:16079. <https://doi.org/10.1038/srep16079>.
20. Kohonen T. Self-Organizing Maps. Springer, Berlin Heidelberg; 2001.
21. Weishaar JL, Aiken GR, Bergamaschi BA, Fram MS, Fujii R, Mopper K. Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* 2003;37:4702-4708. <https://doi.org/10.1021/es030360x>.
22. Helms JR, Stubbins A, Ritchie JD, Minor EC, Kieber DJ, Mopper K. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* 2008;53:955-969. <https://doi.org/10.4319/lo.2008.53.3.0955>.
23. Chin YP, Aiken G, O'Loughlin E. Molecular-weight, polydispersity, and spectroscopic properties of aquatic humic substances. *Environ. Sci. Technol.* 1994;28:1853-1858. <https://doi.org/10.1021/es00060a015>.
24. Hansen AM, Kraus TEC, Pellerin BA, Fleck JA, Downing BD, Bergamaschi BA. Optical properties of dissolved organic matter (DOM): effects of biological and photolytic degradation. *Limnol. Oceanogr.* 2016;61:1015-1032. <https://doi.org/10.1002/lno.10270>.
25. McKnight DM, Boyer EW, Westerhoff PK, Doran PT, Kulbe T, Andersen DT. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* 2001;46:38-48. <https://doi.org/10.4319/lo.2001.46.1.0038>.
26. Ohno T. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. *Environ. Sci. Technol.* 2002;36:742-746. <https://doi.org/10.1021/es0155276>.
27. 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. [https://doi.org/10.1016/S0146-6380\(00\)00124-8](https://doi.org/10.1016/S0146-6380(00)00124-8).
28. Bro R. PARAFAC: tutorial and applications. *Chemom. Intell. Lab. Syst.* 1997;38:149-171. [https://doi.org/10.1016/S0169-7439\(97\)00032-4](https://doi.org/10.1016/S0169-7439(97)00032-4).
29. Stedmon CA, Markager S, Bro R. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* 2003;82:239-254. [https://doi.org/10.1016/S0304-4203\(03\)00072-0](https://doi.org/10.1016/S0304-4203(03)00072-0).
30. Cory RM, McKnight DM. Fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter. *Environ. Sci. Technol.* 2005;39:8142-8149. <https://doi.org/10.1021/es0506962>.
31. Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. Self-organizing map in Matlab: the SOM Toolbox in Proceedings of the Matlab DSP Conference. 1999. Espoo, Finland.
32. South Korea Ministry of Environment. Strategies for Effective Control and Treatment of Urban Stormwater in Wet-Weather Conditions; 2018. Available from: http://www.me.go.kr/home/web/policy_data/read.do?pagerOffset=0&maxPageItems=10&maxIndexPages=10&searchKey=&searchValue=&menuId=10259&orgCd=&condition.deleteYn=N&seq=7360.
33. Huber WC, Dickinson RE. Storm Water Management Model Version 4, User's Manual. 1988, EPA/600/3-88/001a, United States Environmental Protection Agency: Athens, GA. Available from: <http://www.dynsystm.com/netstorm/docs/swmm4manuals.pdf>.
34. Vallabhaneni S, Chan C, Burgess E. Computer tools for sanitary sewer system capacity analysis and planning, EPA/600/R-07/111. 2011. Available from: https://cfpub.epa.gov/si/si_public_record_Report.cfm?Lab=NRMRL&dirEntryID=184303.
35. Nam SN, Amy G. Differentiation of wastewater effluent organic matter (EfOM) from natural organic matter (NOM) using multiple analytical techniques. *Water Sci. Technol.* 2008;57:1009-1015. <https://doi.org/10.2166/wst.2008.165>.

36. Korak JA, Wert EC, Rosario-Ortiz FL. Evaluating fluorescence spectroscopy as a tool to characterize cyanobacteria intracellular organic matter upon simulated release and oxidation in natural water. *Water Res.* 2015;68:432-443. <https://doi.org/10.1016/j.watres.2014.09.046>.
37. Dong MM, Rosario-Ortiz FL. Photochemical Formation of Hydroxyl Radical from Effluent Organic Matter. *Environ. Sci. Technol.* 2012;46:3788-3794. <https://doi.org/10.1021/es402491t>.
38. Carpenter KD, Kraus TEC, Goldman JH et al. Sources and characteristics of organic matter in the Clackamas River, Oregon, related to the formation of disinfection by-products in treated drinking water. U.S. Geological Survey Scientific Investigations Report 2013-5001; 2013.
39. Fleck JA, Gill G, Bergamaschi BA, Kraus EC, Downing BD, Alpers CN. Concurrent photolytic degradation of aqueous methylmercury and dissolved organic matter. *Sci. Total Environ.* 2014;484:263-275. <https://doi.org/10.1016/j.scitotenv.2013.03.107>.
40. Jaffé R, McKnight D, Maie N, Cory R, McDowell WH, Campbell JL. Spatial and temporal variations in DOM composition in ecosystems: The importance of long-term monitoring of optical properties. *J. Geophys. Res. Biogeosci.* 2008;113:1-15. <https://doi.org/10.1029/2008JG000683>.
41. Wilson HF, Xenopoulos MA. Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nat. Geosci.* 2009;2:37-41. <https://doi.org/10.1038/ngeo391>.
42. Murphy KR, Stedmon CA, Waite TD, Ruiz GM. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Mar. Chem.* 2008;18:40-58. <https://doi.org/10.1016/j.marchem.2007.10.003>.
43. Asmala E, Massicotte P, Carstensen J. Identification of dissolved organic matter size components in freshwater and marine environments. *Limnol. Oceanogr.* 2021;66(4):1381-1393. <https://doi.org/10.1002/lno.11692>.
44. Nam SN. Characterization and differentiation of wastewater effluent organic matter (EfOM) versus drinking water natural organic matter (NOM): implications for indirect potable reuse [dissertation]. Boulder: Univ. of Colorado at Boulder; 2011.
45. Cawley KM, Ding Y, Fourqurean J, Jaffe R. Characterising the sources and fate of dissolved organic matter in Shark Bay, Australia: a preliminary study using optical properties and stable carbon isotopes. *Mar. Freshw. Res.* 2012;63:1098-1107. <https://doi.org/10.1071/MF12028>.
46. Baghoth SA, Sharma SK, Amy GL. Tracking natural organic matter (NOM) in a drinking water treatment plant using fluorescence excitation-emission matrices and PARAFAC. *Water Res.* 2011;45:797-809. <https://doi.org/10.1016/j.watres.2010.09.005>.
47. Murphy KRM, Butler KD, Spencer RGM, Stedmon CA, Boehme JR, Aiken GR. Measurement of dissolved organic matter fluorescence in aquatic environments: an interlaboratory comparison. *Environ. Sci. Technol.* 2010;44:9405-9412. <https://doi.org/10.1021/es102362t>.
48. Coble PG, Lead J, Baker A, Reynolds DM, Spencer RGM. Aquatic Organic Matter Fluorescence. Cambridge University Press; 2014.
49. Stedmon CA, Markager S, Tranvik L, Kronberg L, Slätis T, Martinsen W. Photochemical production of ammonium and transformation of dissolved organic matter in the Baltic Sea. *Mar. Chem.* 2007;104(3):227-240. <https://doi.org/10.1016/j.marchem.2006.11.005>.
50. Goncalves-Araujo R, Stedmon CA, Heim B et al. From Fresh to Marine Waters: Characterization and fate of dissolved organic matter in the Lena River Delta region, Siberia. *Front. Mar. Sci.* 2015;2:23-35. <https://doi.org/10.3389/fmars.2015.00108>.