# Source-tracing of industrial and municipal wastewater effluent in river water via fluorescence fingerprinting

Sandra Peer [a,*], Anastassia Vybornova [a,b], Zdravka Saracevic [a], Jörg Krampe [a], Ottavia Zoboli [a]

[a] *Institute for Water Quality and Resource Management, TU Wien, Karlsplatz 13/226, 1040 Vienna, Austria*
[b] *IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen, Denmark*

## HIGHLIGHTS

- Fluorescence fingerprinting identifies unique DOM compositions in water samples.
- Seven PARAFAC components characterize WWTP effluents from various industries.
- PCA clusters WWTP effluents and source-traces them in the receiving river.
- At low flow WWTP emissions dominate, at high flow diffuse emissions mask WWTPs.
- A technique suitable for implementing complex fluorescence monitoring in rivers.

## GRAPHICAL ABSTRACT
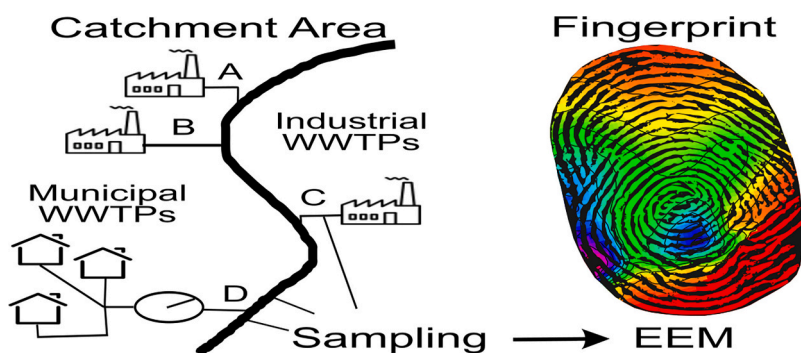


## ARTICLE INFO

## ABSTRACT

Fluorescence fingerprinting is a technique to uniquely characterize water samples based on their distinct composition of dissolved organic matter (DOM) measured via 3D fluorescence spectroscopy. It is an effective tool for monitoring the chemical composition of various water systems. This study examines a river affected by several municipal and industrial wastewater treatment plant (WWTP) effluents and aims to source-tracing them via fluorescence fingerprints based on parallel factor analysis (PARAFAC) components. Additional principal component analysis (PCA) clusters the WWTP effluents according to similarity. The results yield seven PARAFAC components characterizing the WWTP effluents. Considering the ratios among the components, these distinct fluorescence fingerprints are attributable to the studied industrial sectors: leather industry, meat processing, electronics industry, and municipal wastewater treatment. Furthermore, the fluorescence signal of the receiving river is examined by PCA and assessment of flow-weighted fluorescence intensities for source-tracing the fingerprints of the WWTP effluents. An analysis of the contribution of each WWTP effluent shows that during low flow, the fluorescence signal in the river is dominated by WWTP emissions. In contrast, during high flow events, the impact of WWTP emissions is masked by diffuse emissions. The techniques presented in this study have the potential to define generalizable fluorescence fingerprints for WWTP effluents of various industrial sectors and source-trace them in the receiving river. This approach represents a step closer to implementing complex

fluorescence monitoring tools in rivers, tracing the impact of municipal and industrial WWTP effluents on riverine OM.

## 1. Introduction

3D fluorescence spectroscopy has become well known for its easy and rapid water quality assessment by detecting dissolved organic matter (DOM) (Coble, 1996; Coble et al., 2014). Its diverse range of applications stretches from drinking water to natural waters and wastewater (Hudson et al., 2007; Henderson et al., 2009; Carstea et al., 2016). Precisely for this reason, fluorescence spectroscopy is gaining increasing importance in environmental monitoring and protection (Shen et al., 2021).

Fluorescence fingerprinting offers a unique perspective in this regard. Depending on the distinct DOM composition of each water sample, a specific fingerprint is obtained. This fingerprint is determined by the intensity of the fluorescence signal in different DOM areas and is designed to distinguish the corresponding water source (Rodríguez-Vidal et al., 2020). This is highly beneficial, for instance, for monitoring wastewater treatment processes (Carstea et al., 2016). Work has been ongoing recently to identify generalized fluorescence fingerprints for WWTP effluents, although few studies have been published as yet (Liu et al., 2019). While municipal WWTP effluents have already been discussed in detail, industrial WWTP effluents have proven more challenging to obtain a fluorescence fingerprint due to their more complex composition. Nevertheless, the potential has been clearly demonstrated (Rodríguez-Vidal et al., 2020). In some sectors, identifying specific fluorescence fingerprints may be more straightforward than in others. For example, effluents from food processing plants show very consistent fingerprints, similar to municipal WWTP effluents (Rodríguez-Vidal et al., 2020). For metal plating plants, on the contrary, differences in production processes lead to varying fluorescence fingerprints (Shen et al., 2021). Generally, more research is needed to cover the wide range of industrial sectors and provide a collection of specific fingerprints (Rodríguez-Vidal et al., 2020, 2022).

Opinions diverge as to which method is best for creating a fluorescence fingerprint. Rodríguez-Vidal et al. (2020) attempt to define fingerprints using peak picking and fluorescence indices, though with no clear recommendation for either of these methods. Another promising approach is extracting components using parallel factor analysis (PARAFAC), a statistical method for dimensional reduction of high-dimensional excitation-emission matrices (EEM) (Murphy et al., 2013). From these, fluorescence fingerprints may be determined through advanced analysis (Yang et al., 2015).

Beyond the definition of the fluorescence fingerprint of WWTP effluents, novel perspectives for environmental monitoring are emerging. Numerous studies have shown that fluorescence spectroscopy is an effective tool for monitoring changes in the chemical composition of various water bodies, as is the case for merging source rivers or WWTP effluent discharge (Yan et al., 2000; Goldman et al., 2012; Liu et al., 2019; Rodríguez-Vidal et al., 2020, 2022; Shen et al., 2021; Matos et al., 2022). Previous studies already demonstrated that the specific fluorescence fingerprint of WWTP effluents remains recognizable even after intermixing in the receiving river (Baker, 2001). In addition, it is even possible to monitor their effect and behavior along streams and rivers (Sgroi et al., 2017).

In light of these findings, fluorescence fingerprinting proved particularly suitable for tracing changes in municipal WWTP emissions in natural waters, especially regarding the protein-like fluorescence peak (Cawley et al., 2012). Residual PARAFAC components have been identified by comparing them with PARAFAC components from global models of unpolluted rivers, which represent the specific anthropogenic source of DOM where present (Cawley et al., 2012; Liu et al., 2019). For single point emissions, the WWTP effluent signature can be inferred

directly by matching upstream and downstream river samples (Baker, 2001). Complex fluorescence fingerprints may be monitored using system-specific PARAFAC models and, in certain scenarios, may even be associated with specific DOM sources (Cawley et al., 2012), provided there is sufficient variation in fluorescence properties (Li et al., 2014). However, attention must always be paid to the fluorescence signal of the water body itself since it cannot be concluded with absolute certainty that a specific pollution source has occurred due to interference of the fluorescence signals (Liu et al., 2019).

Even though numerous efforts are aimed at fingerprinting WWTP effluents in rivers, these studies focus either on single WWTPs or on several WWTPs in the same industrial sector (Baker, 2001; Cawley et al., 2012; Liu et al., 2019). However, the approach becomes more difficult, but also more promising, when several WWTPs from different industrial sectors coalesce along a river (Yang et al., 2015). Precisely, in this case, unique and generalizable fluorescence fingerprints could offer major benefits, such as monitoring standard conditions and detecting anomalies or violations of regulatory criteria (Baker, 2001; Yang et al., 2015). The ability to rapidly detect which WWTPs deviate from the standard fingerprint typical for their sector, for example, in the event of a technical failure, is invaluable (Shen et al., 2021). As a result, adequate measures can be implemented at the water body and within the technical area of the WWTPs to protect the environment as effectively as possible (Cawley et al., 2012; Yang et al., 2015; Liu et al., 2019).

For this reason, the present study examines a river affected by municipal and industrial WWTP effluents of various sectors. The aim is to obtain a fluorescence fingerprint for each WWTP effluent based on the PARAFAC components. These fluorescence fingerprints are also evaluated across the sampling period to reveal seasonal or production-related cycles. Moreover, the fingerprints are examined in terms of similarity and grouped into clusters showing similar fluorescence properties, enabling the identification of similarities and differences between the fluorescence fingerprints of the various kinds of WWTP effluents. Ideally, this could generalize the fluorescence fingerprints beyond the scope of this study. Furthermore, the longitudinal evolution of the fluorescence signal along the river and tracing of the fluorescence signal of the WWTP effluents in the river by two complementary methods, namely by PCA as a well-established statistical method and assessment of flow-weighted fluorescence intensities as a novel approach in this context, is shown. Consequently, the potential and limitations of fluorescence fingerprints in establishing associations with the fluorescence signal of the river after intermixing are discussed. This approach constitutes an important step towards finding a tool for monitoring rivers and their discharges, which recognizes anomalies early and provides indications regarding potential pollution sources.

## 2. Materials and methods

### 2.1. Sampling sites

Between November 2020 and February 2022, monthly river water samples (a total of 14 samples) were taken at an online monitoring station at a lowland river with a catchment area of 1,009 km$^2$ with mixed land use. Approximately 3 % of the river's total mean flow of 9.9 m$^3$ s$^{-1}$ is contributed by the effluents of municipal and industrial WWTPs. An online monitoring station is situated further downstream of all known emitters. It serves, among others, the purpose of long-term river monitoring of water quality parameters and supervision of municipal and industrial WWTP emissions. All samples were taken under dry weather except for the sample in December 2020.

Additionally, a total of 80 corresponding effluent samples were

collected from a selection of six different WWTPs that contribute their effluent into the river (the letters representing the WWTPs indicate their location order along the river course from upstream to downstream): (1) a municipal WWTP with a design capacity of $32,000$ population equivalents (PE) based on 120 g chemical oxygen demand per PE and day ($COD_{120}$) (WWTP B), (2) an industrial WWTP of a meat processing mill with 25,900 PE (WWTP C), (3) an industrial WWTP of an electronics manufacturer (printed circuit boards and substrates for semiconductors) with chemical treatment exclusively for process water (WWTP E) and (4–6) three industrial WWTPs of leather mills with $50,000$ PE (WWTP A), 232,000 PE (WWTP D), and 60,000 PE (WWTP F), respectively. Table 1 gives an overview of the hydraulic flow and standard water quality parameters of all WWTPs during the sampling period.

When selecting these WWTPs, the aim was to cover the diversity of all discharging WWTPs as comprehensively as possible. The selected municipal WWTP is the largest in the catchment and accounts for 30 % of the water discharge from municipal WWTPs and 25 % from all WWTPs. Due to the similarity of municipal wastewater, this plant is considered representative of this input type. Altogether, the six selected WWTPs account for 40 % of the overall water discharge from WWTP effluents. Considering all industrial WWTPs, the five WWTPs included in this study contribute 84 % of the industrial water discharge. All three WWTPs for the leather industry in the catchment have been included in the study. They contribute to about 73 % of the wastewater discharges from industrial WWTPs and only about 13 % to the discharges from all WWTPs. On the other hand, these three plants from the leather industry dominate the organic inputs into the river. Nearly all industrial sectors present at the river are represented in the emitter samples. Any missing share consists of two additional WWTPs, a meat processing mill with a lower design capacity than the one included, and a fruit processing mill, which could not be incorporated into this study for logistic reasons. In this respect, the selection is highly representative of the overall wastewater loading of the considered river.

In order to analyze in detail each effluent's effect on the river's DOM as a result of mixing, additional longitudinal sampling was conducted at ten sampling points along the river on June 17, 2021. For this purpose, a sample was taken before the first discharge at WWTP A and again after each successive WWTP sampled. Proper distances were chosen to ensure the complete intermixing of the effluents with the river water. The two concluding sampling sites are the above-described online monitoring station and a site located approximately 10.2 km downstream of the monitoring station to investigate the further alteration of the DOM without any additional influences. Fig. 1 provides a schematic map of the study area, including the sampled WWTPs' location and all sampling points along the river.

## 2.2. Analytical methods

Standard water quality analysis of the samples comprised of total organic carbon (TOC) (DIN EN 1484), dissolved organic carbon (DOC) (DIN EN 1484), chemical oxygen demand (COD) (DIN 38409-43 and DIN ISO 15705), ammonia nitrogen ($NH_4$-N) (DIN EN ISO 11732), nitrate nitrogen ($NO_3$-N) (DIN EN ISO 13395), total nitrogen (TN) (DIN EN ISO 11905-1), and total phosphorus (TP) (DIN EN ISO 6878).

Samples were immediately filtered with a 0.45 μm membrane filter

and stored no longer than two days under cool (4 °C) and dark conditions, following the recommended standard protocol established by Peer et al. (2023). Prior to the spectroscopic measurement, samples were allowed to warm up to room temperature (~ 20 °C). Additionally, river water samples were four-fold diluted, samples of WWTPs B, C, and E were five-fold diluted, and samples of WWTPs A, D, and F were 40-fold diluted with Milli-Q in order to record excitation-emission matrices (EEM) using a HORIBA Scientific Aqualog® spectrofluorometer equipped with a Xenon lamp. The inner filter effect was reduced by choosing each dilution so that a maximum absorbance of 1.5 was not exceeded. The measurement procedure and correction steps followed precisely the methods outlined in Peer et al. (2022) with an integration time of 2 s for all samples and a final transformation of the resulting EEM into Raman units [R.U.] (Lawaetz and Stedmon, 2009).

### 2.3. Methods for determining, clustering and tracing fluorescence fingerprints

The statistical software R (R Core Team, 2021) was used throughout all analyses. A Parallel Factor Analysis (PARAFAC) with non-negativity constraints using the procedure outlined in the package *staRdom* (Pucher et al., 2019) was applied in order to extract four to eight latent components from the trilinear multi-way data structure of the 80 analyzed EEMs of the emitters. For determining the practical significance and the number of components, the explained variance, the residual sum of squares, and the core consistency score (CORCONDIA) (Bro and Kiers, 2003) were calculated for the PARAFAC models comprising four to eight components. Validation of the final seven-component model via multiple random initializations (Harshman and Lundy, 1994), split-half and residual analysis (Murphy et al., 2013), as well as the transformation of the resulting components to Raman Units (Pucher et al., 2019) representing the relative DOM concentration, was performed as described in Peer et al. (2022). Additionally, the PARAFAC sample loadings for samples not involved in model building for the emitter PARAFAC model were calculated using the corresponding function from the *staRdom* package (Pucher et al., 2019). More specifically, 21 river samples were examined separately to evaluate the distribution of the emitter PARAFAC components in the river water. Furthermore, the emitter PARAFAC model was applied to ten longitudinal river samples. In both cases, a residual analysis was also performed to ensure no residual signal remained undetected by the emitter PARAFAC model.

To determine a fingerprint of the emitter samples relative to each other, the percent distribution of PARAFAC components relative to the total DOM per sample (%C1-%C7), called the chemical composition indices according to Yang et al. (2014, 2015), was calculated. Although PARAFAC components are independent of each other by definition, the percentage distribution generates a statistical dependency between these indices as they sum to 100 % for each sample. Such a dependence creates considerable limitations in the further analysis and interpretation of the results, which can be avoided by an additional dimension reduction on the emitter samples via principal component analysis (Pearson, 1901) (PCA) using the function *prcomp* from the *stats* package (R Core Team, 2021). Prior to PCA, the indices were zero-centered and scaled to unit variance to ensure that the analysis was not affected by differences in the scale. The resulting set of principal components

**Table 1**
Mean of hydraulic flow and standard water quality parameters on sampling days for each WWTP. Values in parentheses, when present, indicate the standard deviation. The letters representing the WWTPs are given according to the location from upstream to downstream.

| WWTP | Sector | Q [m³ d⁻¹] | TOC [mg L⁻¹] | DOC [mg L⁻¹] | COD [mg L⁻¹] | NH₄-N [mg L⁻¹] | NO₃-N [mg L⁻¹] | TN [mg L⁻¹] | TP [mg L⁻¹] |
|---|---|---|---|---|---|---|---|---|---|
| A | Leather | 1,278 (139) | 48.5 (13.4) | 40 (12.7) | 117 (40.9) | 1.1 (1.6) | 25.6 (15.9) | 39.4 (15.5) | 0.4 (0.1) |
| B | Municipal | 6,879 (1,861) | 6.75 (1) | 5.26 (0.8) | 17.8 (6.8) | < 0.1 | 6.2 (3.7) | 7.8 (4.4) | 0.4 (0.1) |
| C | Meat | 398 (10.7) | 8.84 (2.6) | 5.76 (1.7) | 20.7 (6.2) | 3.8 (9.9) | 4.6 (2.8) | 10.8 (10.6) | 0.2 (0.1) |
| D | Leather | 998 (476) | 23.9 (10.7) | 22 (10.6) | 67 (33.2) | < 0.1 | 52.6 (32) | 52.1 (36) | < 0.2 |
| E | Electronics | 162 (18.1) | 32.4 (5.8) | 31 (6) | 57.7 (14.5) | 0.2 (0.2) | 0.1 (0.1) | 1.6 (0.4) | 0.3 (0.2) |
| F | Leather | 1,268 (144) | 47.6 (6.3) | 43.3 (5.3) | 137 (14.8) | 0.2 (0.1) | 53.8 (15.1) | 54.1 (18.8) | 0.7 (0.6) |
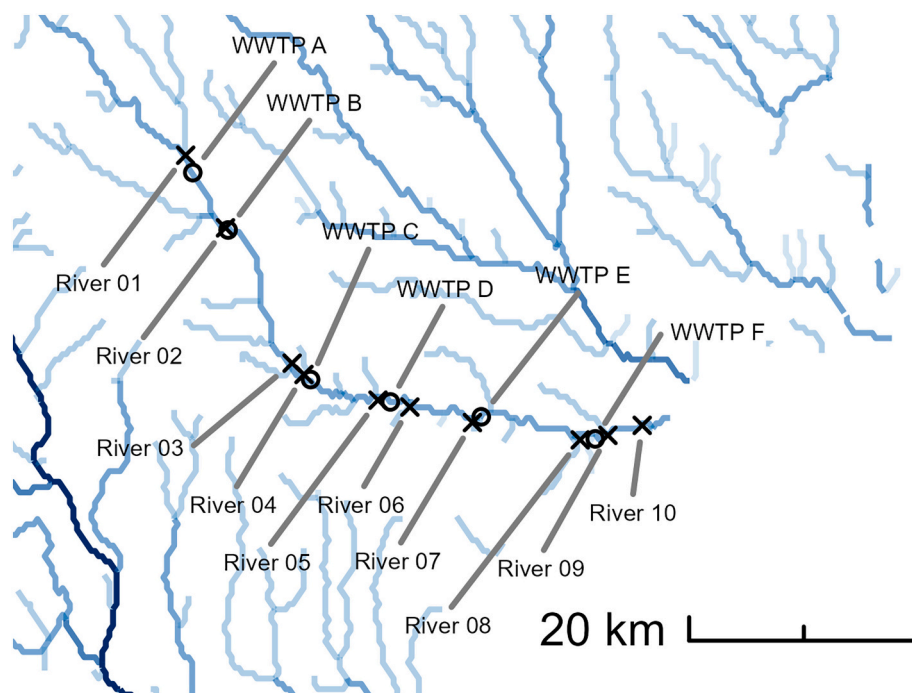
**Fig. 1.** Map of the study area, including the location of the WWTPs and sampling points along the river.

explains the most significant amount of variance in the chemical composition indices. Beyond that, the principal components of the river samples were projected onto the two-dimensional space created via PCA to see how these river samples fit into the clusters obtained based on the emitter samples. This projection allows to assess the resemblance of the river and emitter samples, reveals similarities in the individual PARAFAC components, and provides potential associations between the fluorescence fingerprints of the emitters and the river.

Calculations of substance flows/loads from substance concentration and hydraulic flow expressed as mass per time are the basis of any mass balance calculations in water resource systems research to identify impacts from diffuse or point emissions on receiving water bodies. The Water Framework Directive (2000/60/EC) of the European Commission (EC) suggests a tiered approach for relating mass balance calculations to river pollution (EC, 2012). Fluorescence intensity does not directly represent concentrations but correlates with DOM concentrations represented by PARAFAC components. In order to combine fluorescent techniques with requirements from water quality management, the riverine approach (tier two) is innovatively utilized, and the concept of "flow-weighted fluorescence intensity" for quantitatively tracing fluorescence fingerprints of emitters in rivers is implemented. To calculate the flow-weighted fluorescence intensity, the intensities of the PARAFAC components C1 to C7 (R.U.) of each emitter are multiplied by the specific hydraulic flow of the sampling day ($m^3 d^{-1}$). Consequently, the sum of all emitters is divided by the river flow of the corresponding day at the online monitoring station, leading to a theoretical fluorescence intensity that can be addressed to emitters if persistent behavior of fluorescent DOM is assumed. For each PARAFAC component, the calculated instream intensity stemming from emitters is compared to the corresponding monitored fluorescence intensities in the river at the online monitoring station, indicating to which extent the fluorescence intensity of a component can be explained by emitters and which share is not explained and may be attributed to diffuse inputs according to the riverine approach (tier two) (EC, 2012).

## 3. Results and discussion

### 3.1. PARAFAC components to define fluorescence fingerprints of the WWTP effluents

As a first visualization, the fluorescence signals of the six WWTP samples are presented through an exemplary contour plot for each selected sample in Fig. 2, representing the typical fluorescence signal at dry weather conditions. The contour plots map all combinations of the recorded excitation and emission wavelengths in a grid structure on the X- and Y-axes. The color scale of the plot indicates the fluorescence intensity in Raman Units (R.U.), with darker shades indicating higher intensities.

WWTP B has a typical fluorescence signal for a municipal WWTP with two characteristic peaks, the first one at an excitation of about 300 nm and an emission of about 341 nm and the second one at an excitation of about 225 nm and an emission of about 346 nm (Baker, 2001). While the signal of industrial WWTP C (meat processing) appears very similar, both peaks are located at a slightly higher emission wavelength of about 406 nm. These peaks also occur in WWTP B but are only minor peaks with a notably lower intensity. WWTP E (electronics manufacturer) reveals the lowest signal of all the WWTPs analyzed, with the main peak at an excitation of about 228 nm and an emission of about 305 nm. Generally, WWTPs A, D, and F (industrial WWTPs of leather mills) show significantly higher fluorescence signals than the WWTPs mentioned earlier. While the shape of the fluorescence signal is roughly the same as that of industrial WWTP E, the primary peak of WWTPs A, D, and F is found at a higher emission wavelength of approximately 346 nm.

Relevant indicators were compared across models with four to eight components to determine the number of components in the final PARAFAC model (Table 2). As the explained variance is very high for all models considered, this indicates a good fit of the models to the samples. The residual sum of squares reveals a substantial decline once seven components are included, followed by a more moderate decline when further components are added. The CORCONDIA scores are generally very low, especially for the models with six or more components. However, this is to be expected for models with more than five components, as this metric tends to prevent over-fitting (Murphy et al.,
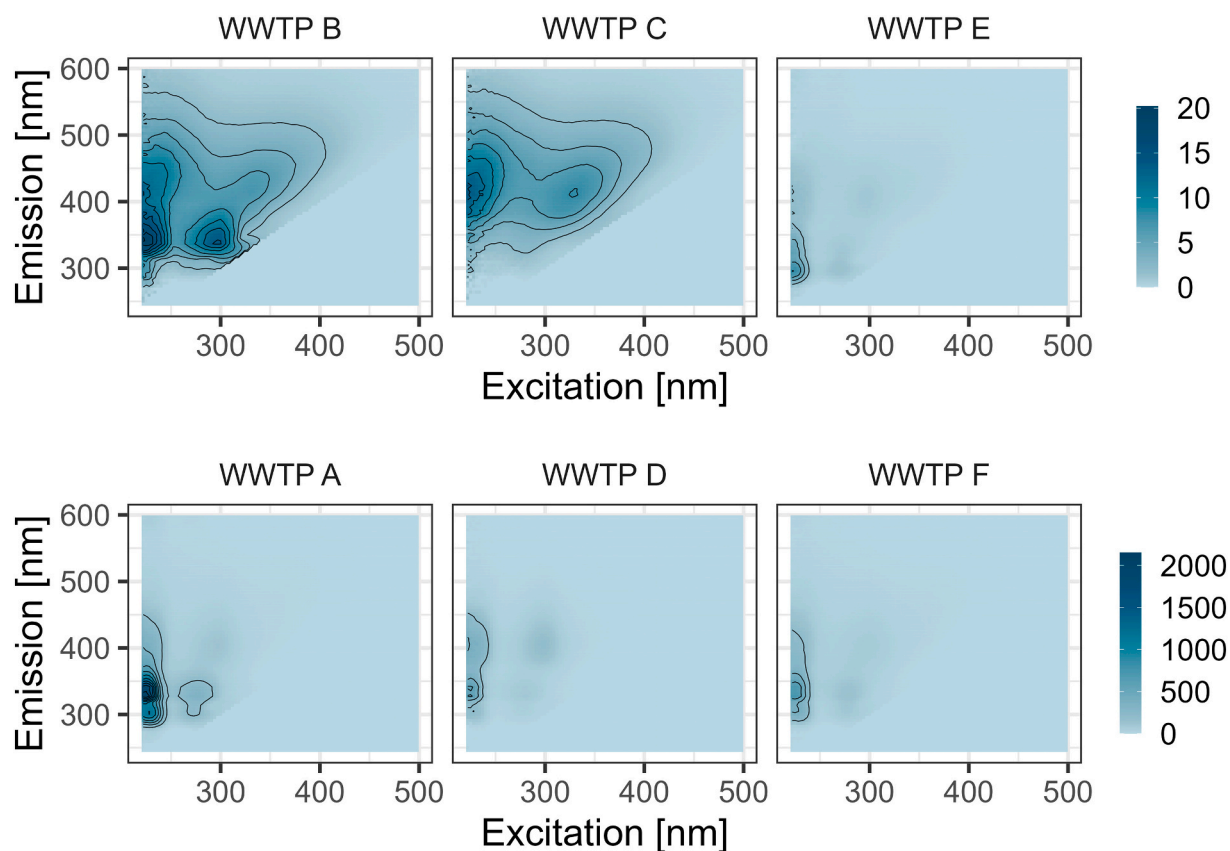
**Fig. 2.** Exemplary contour plots of the sampled effluents of the six sampled WWTPs emitting into the monitored river. The color scale indicates the fluorescence intensity [R.U.].

**Table 2**
Explained variance, residual sum of squares, and CORCONDIA scores for PARAFAC models with four to eight components (No.).

| No. | Explained variance [%] | Residual sum of squares | CORCONDIA [%] |
|-----|------------------------|-------------------------|----------------|
| 4 | 96.9 | 371,244.0 | 28.3 |
| 5 | 98.0 | 336,613.7 | 13.7 |
| 6 | 98.5 | 328,005.2 | 0.6 |
| 7 | 98.9 | 115,135.8 | 0.4 |
| 8 | 99.1 | 78,050.9 | 1.0 |

2013). In addition to the statistical indicators, assessing the practical significance of the components is essential. For capturing low levels of fluorophores, which can be decisive for the fingerprint, emphasis is given to the residual analysis. Visual evaluation of the residual plots of all models showed that, finally, the model with seven components could ensure that no residual signal remained undetected. Accordingly, even though the CORCONDIA score is suboptimal, a PARAFAC model with a larger number of components is favored for the research question presented here, as they are superior in interpretability.

The final model of the PARAFAC analysis of all 80 WWTP effluent samples yielded a seven-component model that discriminates different DOM sources. All samples had a leverage below 0.25. A split-half analysis confirmed the validity and stability of the established model (Fig. S1 in the Supplementary material). The contour plots in Fig. 3 show the position of each PARAFAC component, while Table 3 gives their exact primary peak location and description.

Though only a few studies address the fluorescence fingerprint of industrial WWTP effluent, at least some similarities between the presented WWTP effluent model and peaks identified by earlier studies, which also analyzed slaughterhouse WWTP effluents, could be found. A PARAFAC model, for instance, found humic-like (secondary peak of C2/

C3) and protein-like fluorescence (secondary peak of C1) as primary DOM sources (Matos et al., 2022). In contrast, peak picking revealed a tyrosine-like peak in this case (Rodríguez-Vidal et al., 2020).

Due to the WWTP effluents' very different fluorescence signals and intensity, it is necessary to establish a comparable fingerprint to discriminate these effluents. A validated method is the chemical composition index, where each PARAFAC component is expressed as a fraction of the total DOM per sample (%C1-%C7) (Yang et al., 2014, 2015). For each WWTP, this is illustrated in Fig. 4.

When comparing the time-resolved measurements, it is immediately apparent that WWTPs D and F are closely similar in chemical composition throughout the entire monitoring period. Only between November 2021 and January 2022 does the fraction of PARAFAC component C4 decrease in WWTP D while the fraction of PARAFAC component C7 increases slightly. The resemblance is no surprise, as both treat wastewater from leather mills. While WWTP effluents within the same industrial sector tend to be similar, they can exhibit some degree of variation (Yang et al., 2015). This applies explicitly to WWTP A, which shows the highest variance of the WWTPs treating leather mill wastewater across the annual cycle. Between PARAFAC components C4 and C7, there is an interaction in WWTP A, meaning that their fraction always increases or decreases simultaneously. On the other hand, C5 contributes significantly to the fingerprint of WWTP A only in the winter of 2021/22. Especially towards the end of the monitoring period, the fluorescence fingerprints of WWTPs A, D, and F converge further.

Moreover, some similarities are also found between WWTPs B and C, as these two contain considerably higher fractions of the PARAFAC components C2, C3, and C6 than the other WWTPs. That said, the fraction of components C1 and C7 is higher in WWTP B than in WWTP C, while the opposite is true for C2. Combined, the tryptophan-like and humic-like peaks of these two WWTPs maintain an equal ratio, which is
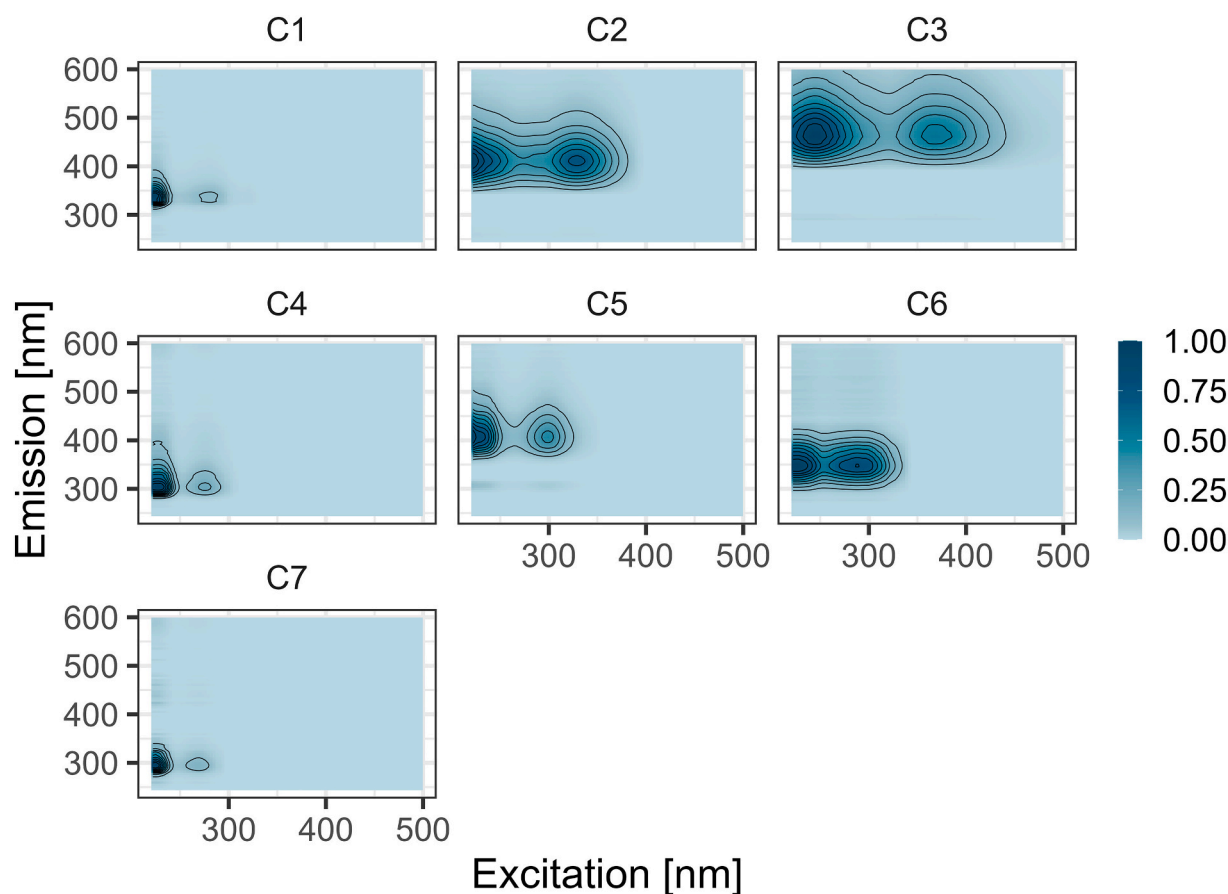
**Fig. 3.** Contour plots of the seven PARAFAC components C1 to C7 normalized to the maximum fluorescence intensity of each component.

**Table 3**
Overview of PARAFAC components found in the samples of the WWTP effluents. Peak labels and descriptions according to Coble (2007) and Bridgeman et al. (2011).

| Component | $\lambda_{ex}/\lambda_{em}$ [nm] | Peak | Description |
|---|---|---|---|
| C1 | 225/341 | $T_2$ | Protein-like (tryptophan) |
| C2 | 222/410 | A | Humic-like (fulvic) |
| C3 | 243/461 | A | Humic-like (fulvic) |
| C4 | 228/305 | $B_2$ | Protein-like (tyrosine) |
| C5 | 222/405 | A | Humic-like (fulvic) |
| C6 | 225/346 | $T_2$ | Protein-like (tryptophan) |
| C7 | 225/296 | $B_2$ | Protein-like (tyrosine) |

particularly expected for municipal WWTPs (Baker, 2001). When directly compared, WWTP E has several unique characteristics. Most evident is the high proportion of PARAFAC component C7. Nevertheless, there are also outliers in this respect, such as in April 2021, where C7 is almost entirely absent, and C1 dominates the fingerprint. In November and December 2021, a slight shift in the fingerprint of WWTP E occurred, resulting in the fractions of all components being more uniformly distributed.

### 3.2. Fluorescence fingerprints to discriminate WWTP effluents

In order to provide statistical evidence for the fingerprints obtained in the previous section, the PARAFAC components were analyzed in greater detail by PCA. As a result, PARAFAC components that comprise the specific fingerprint of each WWTP effluent were identified, and WWTPs with similar fingerprints could be clustered. Both are achieved via a biplot displaying the principal component scores of each sample in a two-dimensional space (Fig. 5). The distance between the samples on the biplot indicates the degree of similarity or dissimilarity between them, resulting in clusters. In addition, the influence of the respective PARAFAC components on the two principal components can be inferred from the vectors displayed. As such, PARAFAC components exhibiting the strongest influence in the direction of the respective cluster result in the distinctive fluorescence fingerprint of the WWTPs associated with this cluster.

The fingerprint of WWTPs A, D, and F (leather mills) is mostly determined by PARAFAC components C1 and C4. PARAFAC components C2, C3, and C6 depict the unique characteristics of the municipal WWTP B and WWTP C (meat processing mill). These two facilities have a similar fluorescence signal and are thus difficult to distinguish by this method. As a result, they form a common cluster. PARAFAC component C5 is only of minor importance in the fingerprint of WWTP E (electronics manufacturer) and WWTPs A, D, and F (leather mills). Despite its minor significance, it clearly distinguishes these two clusters from municipal WWTP B and WWTP C (meat processing). In general, it is quite difficult to establish a fingerprint that discriminates reliably between effluents from municipal wastewater and the food industry as they tend to share tryptophan protein-like peaks (Rodríguez-Vidal et al., 2020). For WWTP E (electronics manufacturer), PARAFAC component C7 determines the fingerprint. Despite its low fluorescence intensity, C7 uniquely differentiates WWTP E from the remaining WWTPs, even if this component shows a fairly large variance in principal component two (PC2). Production-related cycles could explain the variance within the cluster, as has also been observed for WWTPs of metal plating plants, for example (Shen et al., 2021).

As demonstrated, the proposed approach proved highly suitable for characterizing fluorescent fingerprints for municipal and industrial WWTP effluents. Even a limited number of 14 samples or less per WWTP
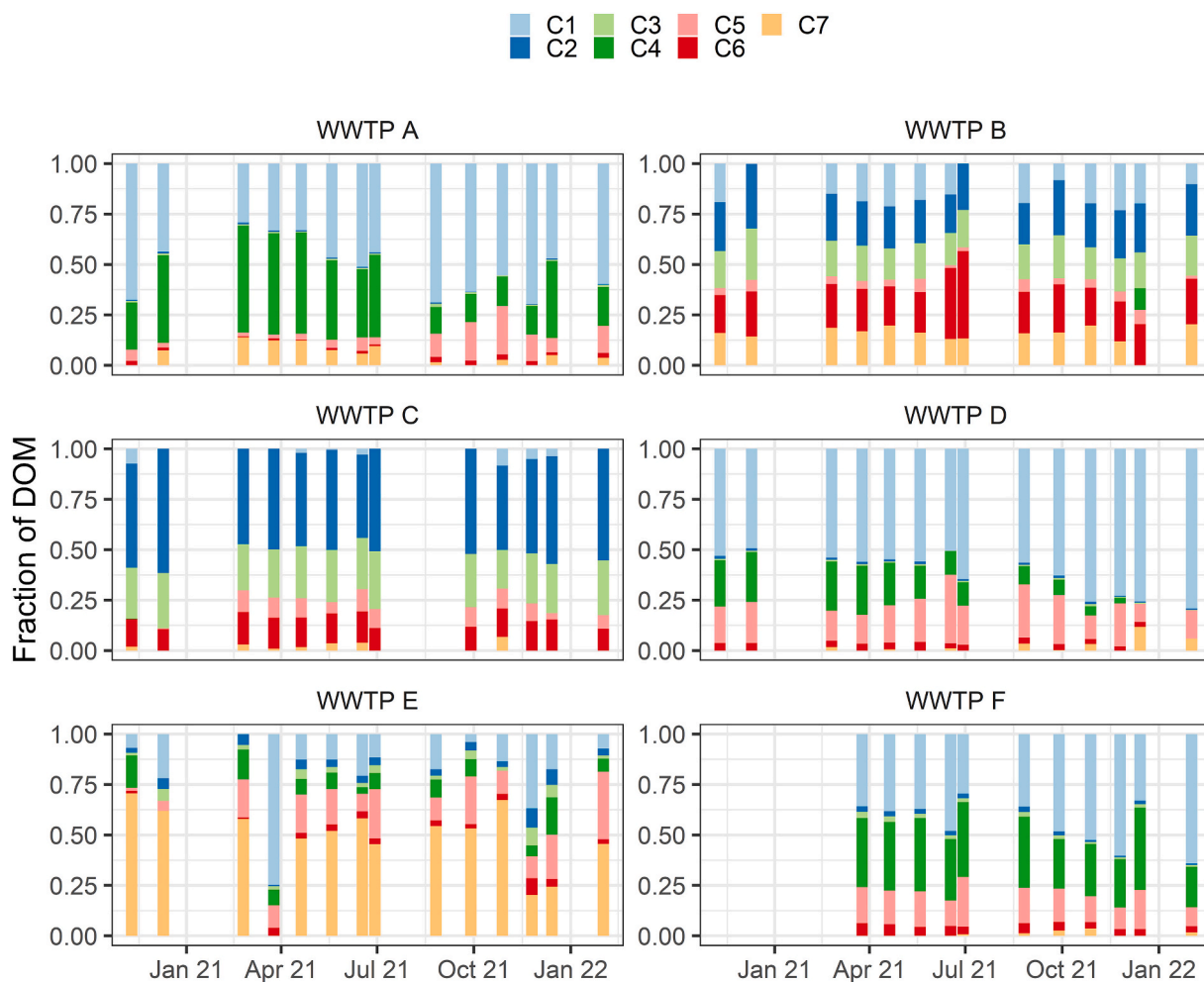
**Fig. 4.** Fluorescence fingerprint of the sampled WWTP effluents represented as the fraction of the extracted PARAFAC components.

can capture variations within industrial sectors and production-related cycles. The standard fluorescent fingerprint might be feasible with an even smaller number of samples, as long as the analyzed sources differ considerably in their DOM composition.

### 3.3. Identification of WWTP fluorescence fingerprints in river water by PCA

Beyond the mere fingerprint of the WWTP effluents, comparing it to the fluorescence signal of the river at the online monitoring station, i.e., downstream of all characterized emitters and after sufficient intermixing, seems particularly interesting in this aspect. Therefore, the PCA scores of the river at the monitoring station at low flow were projected into the biplot in Fig. 5. Flow rates on the respective sampling days ranged from 2 m³ s⁻¹ to 10 m³ s⁻¹. It is evident that PARAFAC components C1 and C4 dominate the fluorescence signal at low flow. As these constitute the fluorescent fingerprints of WWTPs A, D, and F (leather mills), their effluents determine the fluorescence signal of the river at low flow, as expected from their exceptionally high fluorescence intensity and still relevant hydraulic flow. Overall, the fluorescence signal of the river exhibits a much larger variance since it is also influenced by additional factors such as other point and diffuse emissions. With increasing flow, the fluorescence signal of the river becomes less dominated by the PARAFAC components C1 and C4 due to higher dilution of effluents from WWTPs A, D, and F. Instead, it shifts towards the fingerprint of the municipal WWTP B and WWTP C (meat processing) cluster, which C2, C3, and C6 characterize. Such an observation

may be attributed to diffuse emissions at higher flow rates because their fluorescence signal resembles municipal WWTP effluents more than the very specific signal of the other industrial WWTPs' effluents (Rodríguez-Vidal et al., 2020). Moreover, this can be explained by the hydraulic flushing of sewage from the sewer network, which does not affect industrial WWTPs.

The additional projection of the PCA scores of four high flow events into the biplot in Fig. 5 illustrates this effect vividly. An extensive description of these high flow events, including flow patterns and standard water quality parameters like DOC, Cl⁻, PO₄-P, NH₄-N, and TSS concentrations, can be found in Peer et al. (2022). Event G exhibits with 4.7 m³ s⁻¹ to 18.5 m³ s⁻¹ the lowest flow rate of the events considered here, such that the fingerprint is characterized by the PARAFAC components C1 and C4 assigned to WWTPs A, D, and F (leather mills). On the contrary, the signal of event C, the event with the highest flow (annual flood with 21.8 m³ s⁻¹ to 92.1 m³ s⁻¹), is very similar to WWTP C (meat processing). Event B, ranging from 5.7 m³ s⁻¹ to 32.6 m³ s⁻¹, also belongs to this cluster, though it is slightly more similar to municipal WWTP B. However, a strict separation is barely possible due to the overlap. Event E has a flow rate that ranges from 6 m³ s⁻¹ to 19.6 m³ s⁻¹ and is particularly noteworthy. Most samples of this event match the fingerprints of municipal WWTP B and WWTP C (meat processing). Though the amount of this component generally varies, there is one sample with a higher signal in PARAFAC component C7. In the context of a high flow event, this is rather unusual. Nevertheless, considering that this sample marks the end of the event, it is clear by comparing it with the fluorescence signal of the river at low flow that this chemical
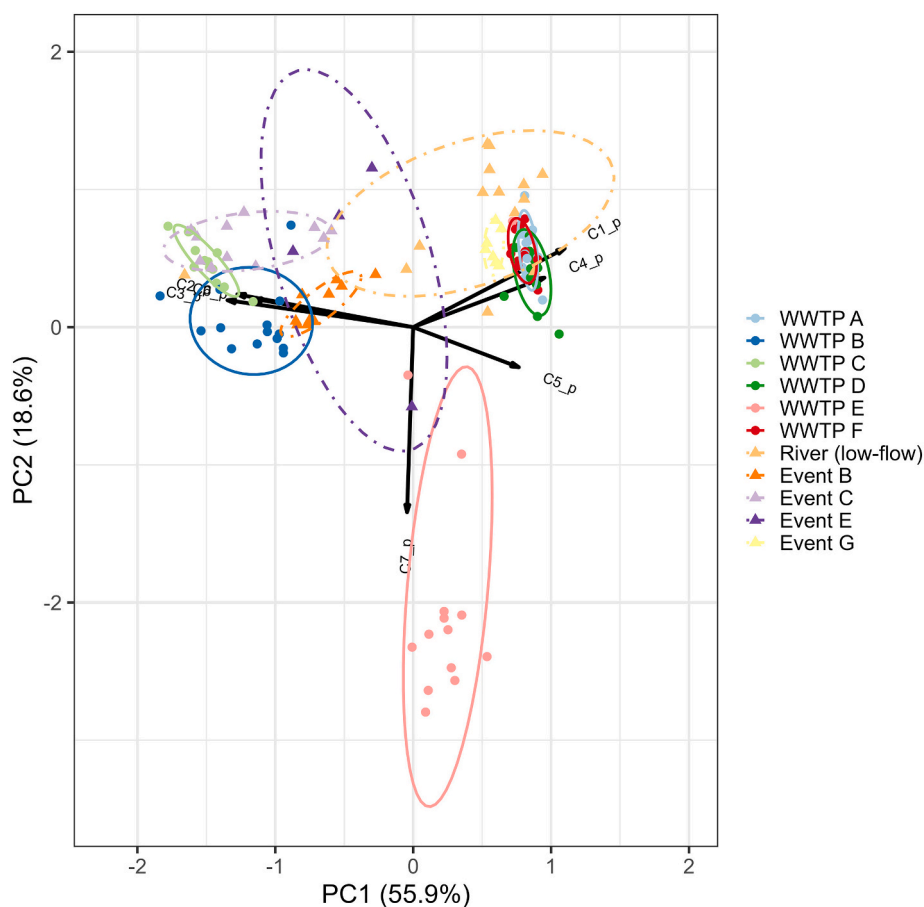
**Fig. 5.** Biplot of principal component 1 (PC1) versus principal component 2 (PC2) from the PCA of the PARAFAC components C1 to C7 of the PARAFAC model based on the WWTP effluent samples. Additionally, the corresponding PCA scores of PARAFAC components extracted from the river samples at low flow conditions and four high flow events were projected into the biplot.

composition may still be expected at 6 m$^3$ s$^{-1}$.

Precisely, PARAFAC component C7 defines the unique fingerprint of WWTP E, which would, therefore, seem particularly suitable for associating it with the fluorescence signal of the river. However, identification is very limited due to its very low fluorescence intensity (max. 5 R.U.), for one, because the fluorescence intensity of the river itself is already much higher (about 50 R.U.), and second, because the fluorescence signal of WWTP E is marginal compared to the WWTP effluents

with a much higher intensity (up to 2000 R.U.). More specifically, the low DOM concentration is diluted and masked during intermixing. An association with the fluorescence signal of the river might be established with a significantly higher DOM concentration at C7 or a less complex constellation.
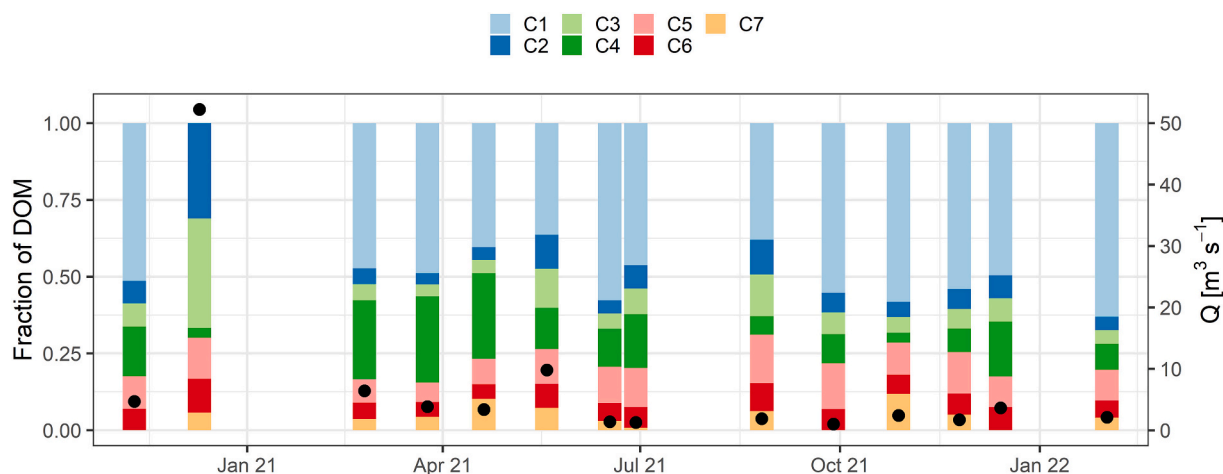


**Fig. 6.** Fraction of the extracted PARAFAC components (C1-C7) of the river samples at the monitoring station, i.e., after intermixing all point emissions. Black dots indicate the hydraulic flow (Q).

### 3.4. Fluorescence fingerprints in and along the river

As indicated in the previous section, the fluorescence signal in the river depends on the hydraulic flow. To illustrate, Fig. 6 provides the chemical composition index of the river at the monitoring station, i.e., after thorough intermixing of all point emissions, for the respective days in the sampling period.

Apparently, the fluorescence signal seen during the high flow event in December 2020 mirrors the fingerprint of municipal WWTP B in Fig. 4, specifically to the fraction of PARAFAC components C2, C3, and C6. These components correspond precisely to the fingerprint of the cluster to which WWTP B is assigned and which overlaps with the signal of diffuse emissions. Given a flow rate of more than 50 m$^3$ s$^{-1}$, the combined sewer overflow might even have been triggered, explaining the close similarity to the chemical composition of municipal WWTP effluent. In the case of low to medium flow (approximately 2 m$^3$ s$^{-1}$ to 10 m$^3$ s$^{-1}$), the chemical composition is relatively stable and does not reveal any significant seasonal patterns.

Most importantly, it was found that the fingerprint of the river matched the fingerprint of diffuse emissions only at high flow. Consequently, as already established by Baker (2001), the additional industrial emissions certainly contribute to the fluorescence fingerprint of the river and, therefore, explain the deviations from the fingerprint of unaffected rivers. A longitudinal sampling campaign in dry weather was conducted to verify this by taking one sample upstream of the first WWTP and after each subsequent WWTP downstream. As a result, the gradual evolution of the river's characteristic fingerprint could be traced in detail (Fig. S2 in the Supplementary material).

The unaffected river sample reveals a well-balanced fluorescence fingerprint, with no single PARAFAC component predominating and a very low fluorescence intensity overall. As early as after the first WWTP, the intensity of the PARAFAC components changed significantly compared to the unaffected river upstream (Baker, 2001). Especially, the intensity of components C1 and C7 rises and prevails distinctly, whereby the expectation that the ratio of tryptophan-like and humic-like components are roughly equal after initiation of a point emission is contradicted (Baker, 2001). Yet not surprisingly, as this is common for municipal WWTPs, WWTP A treats industrial wastewater of a leather mill. While the fluorescence intensity of C1 remains roughly the same farther downstream, the intensity of C7 decreases again after WWTP B. At the same time, the intensity of component C4 also increases, ensuring

the overall fraction of tyrosine-like DOM remains unchanged. As earlier studies suggested, the fraction of humic-like fluorescence increases along the river until it exceeds that of protein-like fluorescence (Wang et al., 2019). The fluorescence fingerprint of the river established at this point is retained up to the last sampling site despite dilution (Baker, 2001). It only shows slight fluctuations in the overall intensity, mainly caused by varying intensities in component C1. Based on the qualitative and quantitative properties of the upstream fluorescence signal, there might be a lack of further changes upon the emission of WWTP effluent. For instance, no differences in the fluorescence fingerprint of the receiving river's upstream and downstream samples were detected for treated slaughterhouse wastewater due to their resemblance to diffuse emissions (Matos et al., 2022). If the two most downstream sampling sites, between which no impact by WWTP effluents takes place, are compared, the high stability of the fluorescent DOM is demonstrated as signals stay highly stable for most of the components (except C1). Slight decreases might be attributed to dilution by diffuse inputs of water.

### 3.5. Flow-weighted fluorescence intensity

Fig. 7 shows the median fraction over the sampling campaign of flow-weighted fluorescence intensity per WWTP as a share of the river's measured fluorescence intensity. Error bars indicate the minimum and maximum fraction per PARAFAC component based on specific days.

The highest contribution is found for WWTP A, especially for C4 and C7, where it provides more than half of the flow-weighted fluorescence intensity from emitters. However, the contribution for C1 and C5 is also noteworthy in this context. Concerning the components C1, C4, C5, and C6, the contribution of WWTPs D and F is also substantial compared to other emitters. Quite the opposite applies to WWTP B, which contributes to components C2, C3, and, to a lesser extent, C7. The share of WWTP C and E is minimal and, therefore, neglectable for the interpretation. Considering the clusters of fluorescence fingerprints as established earlier, it becomes clear that WWTPs A, D, and F (leather mills) cover the most considerable fraction across most components. These components may involve chemical compounds from leather production, which degrade rather slowly.

Solely for components C2 and C3, municipal WWTP B covers a comparably high percentage of flow-weighted fluorescence intensity from emitters. These observations correspond closely to the characteristic components of the fluorescence fingerprints determined earlier. It
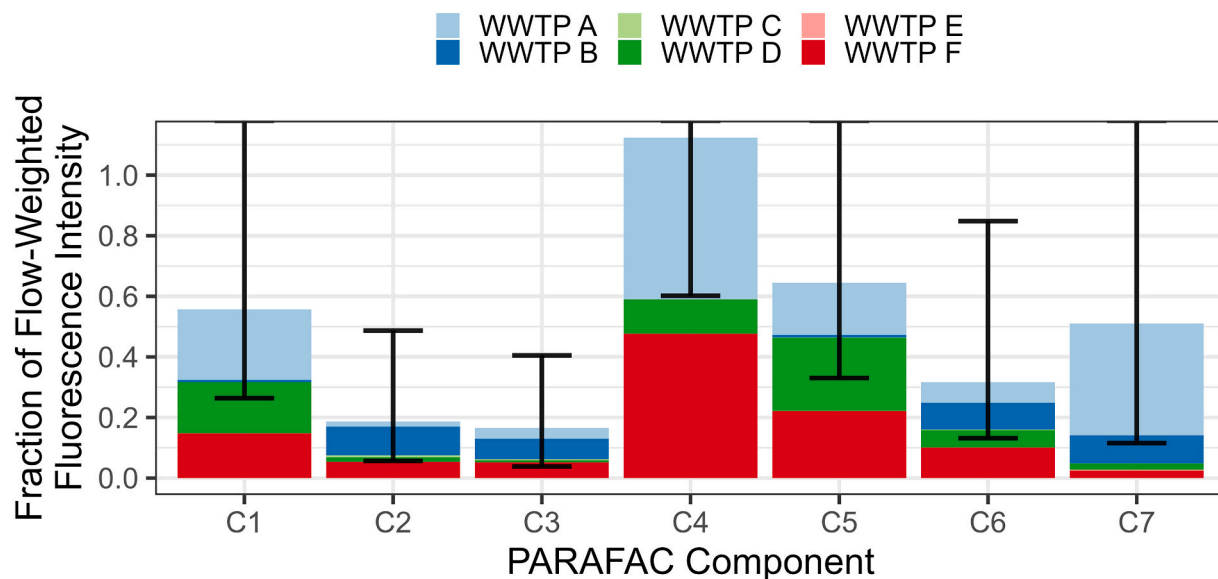


**Fig. 7.** Fraction of flow-weighted fluorescence intensity (median) per WWTP compared to the fluorescence intensity of the river. Error bars indicate the minimum and maximum fraction per PARAFAC component and sampling day. Maxima above 1.1 are truncated.

consists of C1, C4, and partly C5 for the WWTPs of leather mills. For these components, fluorescence intensity from leather mills exceeds, on average, a share of 50 % in the flow-weighted fluorescence intensity in the river. The median fraction for C4 is already 1.1, the maximum for C1 is 1.4, and the maximum for C5 is 1.7, showing limitations as, on several occasions, not all of the flow-weighted fluorescence intensity of the emitters is reflected in the river samples. C2, C3, and, to some extent, C6 and C7 characterize municipal WWTP B. Its impact on the river could have become more visible if the contribution of all municipal WWTPs had been taken into account. Nevertheless, the contribution of municipal WWTP B and WWTP C (meat processing) on average exceeds for none of the PARAFAC components a share of 10 % of the fluorescence in the river, as their fluorescence fingerprint closely matches those of diffuse emissions. Consequently, their contributions cannot be reliably distinguished from the diffuse fluorescent signal of the river, and fingerprints of such WWTPs seem hard to trace in rivers. The missing fraction of the 100 % suggests the presumed effect of the not sampled emitters and the diffuse emissions. Indeed, for components C1, C5, and C7, the monitored WWTPs comprise about half the total amount in the fluorescence signal of the river. Beyond this, the maximum fraction for C7 is undefined, as for several sampling days, this component existed in the WWTP effluents but not in the fluorescence signal of the river. Particularly in this context, the river bed effect must also be taken into account. It states that organic matter is enriched in the sediment so that, for example, a component introduced by an emitter accumulates in the river bed, and only a significantly reduced fraction can be found at the downstream monitoring station. Degradation processes caused by microorganisms in the river may have a similar effect. Part of the fraction not attributable to the emitters might also be explainable by the river bed effect since, depending on the environmental conditions, the previously bound organic matter could also be rereleased into the river water.

These findings indicate that flow-weighted fluorescence intensities have some limitations. Mixture experiments performed with as few as two source waters demonstrated how differences in chemical potential and pH result in non-additivity of the fluorescence signals of the mixture and hence invalidate any simple linear model between fluorescence intensity and concentration (Yan et al., 2000). So far, no generally accepted simple linear model has been established, and given the numerous approaches for establishing source-specific models, it is reasonable to assume that this relationship is considerably more complex (Yan et al., 2000). Without this being fully understood, it is not possible to assess whether the use of flow-weighted fluorescence intensity provides a conclusive interpretation.

## 4. Conclusions and limitations

This study demonstrates the efficacy of fluorescence fingerprinting based on PARAFAC and PCA in characterizing and source-tracing WWTP effluents in the receiving river. By identifying seven distinct PARAFAC components, we have successfully defined distinct fluorescence fingerprints for several industrial sectors, including the leather industry, meat processing, electronics industry, and municipal wastewater treatment. This differentiation is crucial for understanding each sector's unique contributions to the river's overall DOM composition. Further results highlight the significant influence of WWTP effluents on the river's fluorescence signal, particularly during low flow conditions where these emissions dominate. Conversely, during high flow events, the impact of WWTP emissions is masked by diffuse sources.

Limitations regarding this study include contamination of samples or limited comparability of daily composite samples (WWTPs) and grab samples (river). Hence, more validity is given to average results as sudden irregularities have less impact on them. As applying the PARAFAC model for WWTP effluents out-of-sample, i.e., to river water samples, is still experimental (Pucher et al., 2019), it cannot be guaranteed that the PARAFAC components derived from the WWTP model

precisely match the components of a separate model for river samples. Conversely, combining the effluent and river samples would not be appropriate since this yields mixed PARAFAC components, and no specific fluorescence fingerprint for the WWTP effluents could be defined.

Therefore, the techniques presented offer a robust approach to developing generalizable fluorescence fingerprints. The ability to source-trace these fingerprints in the receiving river provides a powerful environmental monitoring and management tool. By advancing the application towards complex fluorescence monitoring, this research contributes to a more precise and effective tracking of municipal and industrial WWTP effluents in riverine systems. In conclusion, this enhances our understanding of the chemical composition of water systems and offers practical solutions for environmental monitoring. Overall, these findings mark an ambitious step towards the source-tracing of fluorescent fingerprints, contributing to the sustainable management of water resources and the protection of aquatic ecosystems.

## CRediT authorship contribution statement

**Sandra Peer:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Anastassia Vybornova:** Writing – review & editing, Software, Investigation, Data curation. **Zdravka Saracevic:** Writing – review & editing, Investigation. **Jörg Krampe:** Writing – review & editing, Resources. **Ottavia Zoboli:** Writing – review & editing, Supervision, Methodology, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.178187.

## Data availability

The authors do not have permission to share data.

## References

Baker, A., 2001. Fluorescence excitation-emission matrix characterization of some sewage-impacted rivers. Environ. Sci. Technol. 35, 948–953. https://doi.org/10.1021/es000177t.
Bridgeman, J., Bieroza, M., Baker, A., 2011. The application of fluorescence spectroscopy to organic matter characterisation in drinking water treatment. Rev. Environ. Sci. Biotechnol. 10, 277. https://doi.org/10.1007/s11157-011-9243-x.
Bro, R., Kiers, H.A.L., 2003. A new efficient method for determining the number of components in parafac models. J. Chemom. 17, 274–286. https://doi.org/10.1002/cem.801.

Carstea, E.M., Bridgeman, J., Baker, A., Reynolds, D.M., 2016. Fluorescence spectroscopy for wastewater monitoring: a review. Water Res. 95, 205–219. https://doi.org/10.1016/j.watres.2016.03.021.

Cawley, K.M., Butler, K.D., Aiken, G.R., Larsen, L.G., Huntington, T.G., McKnight, D.M., 2012. Identifying fluorescent pulp mill effluent in the Gulf of Maine and its watershed. Mar. Pollut. Bull. 64, 1678–1687. https://doi.org/10.1016/j.marpolbul.2012.05.040.

Coble, P.G., 1996. Characterization of marine and terrestrial dom in seawater using excitation-emission matrix spectroscopy. Mar. Chem. 51, 325–346. https://doi.org/10.1016/0304-4203(95)00062-3.

Coble, P.G., 2007. Marine optical biogeochemistry: the chemistry of ocean color. Chem. Rev. 107, 402–418. https://doi.org/10.1021/cr050350+.

Coble, P., Lead, J., Baker, A., Reynolds, D., Spencer, R.G.M., 2014. Aquatic Organic Matter Fluorescence. Cambridge University Press.

European Commission (EC), Directorate-General for Environment, 2012. Technical guidance on the preparation of an inventory of emissions, discharges and losses of priority and priority hazardous substances. In: Guidance Document No 28. European Commission. https://doi.org/10.2779/2764.

Goldman, J.H., Rounds, S.A., Needoba, J.A., 2012. Applications of fluorescence spectroscopy for predicting percent wastewater in an urban stream. Environ. Sci. Technol. 46, 4374–4381. https://doi.org/10.1021/es2041114.

Harshman, R.A., Lundy, M.E., 1994. PARAFAC: parallel factor analysis. Comput. Stat. Data An. 18, 39–72. https://doi.org/10.1016/0167-9473(94)90132-5.

Henderson, R.K., Baker, A., Murphy, K.R., Hambly, A., Stuetz, R.M., Khan, S.J., 2009. Fluorescence as a potential monitoring tool for recycled water systems: a review. Water Res. 43, 863–881. https://doi.org/10.1016/j.watres.2008.11.027.

Hudson, N., Baker, A., Reynolds, D., 2007. Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters—a review. River Res. Appl. 23, 631–649. https://doi.org/10.1002/rra.1005.

Lawaetz, A.J., Stedmon, C.A., 2009. Fluorescence intensity calibration using the Raman scatter peak of water. Appl. Spectrosc. 63, 936–940. https://doi.org/10.1366/000370209788964548.

Li, W.-T., Chen, S.-Y., Xu, Z.-X., Li, Y., Shuang, C.-D., Li, A.-M., 2014. Characterization of dissolved organic matter in municipal wastewater using fluorescence PARAFAC analysis and chromatography multi-excitation/emission scan: a comparative study. Environ. Sci. Technol. 48, 2603–2609. https://doi.org/10.1021/es404624q.

Liu, B., Wu, J., Cheng, C., Tang, J., Khan, M.F.S., Shen, J., 2019. Identification of textile wastewater in water bodies by fluorescence excitation emission matrix-parallel factor analysis and high-performance size exclusion chromatography. Chemosphere 216, 617–623. https://doi.org/10.1016/j.chemosphere.2018.10.154.

Matos, M.C., Tadini, A.M., da Conceição, F.R., Junior, A.M., Menegatti, C.R., Mounier, S., Caires, A.R.L., Nicolodelli, G., 2022. Dissolved organic matter in bovine slaughterhouse wastewater using fluorescence spectroscopy associated with CP/PARAFAC and PCA methods. Appl. Opt., AO 61, 6590–6598. https://doi.org/10.1364/AO.461746.

Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Anal. Methods 5, 6557. https://doi.org/10.1039/c3ay41160e.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2, 559–572. https://doi.org/10.1080/14786440109462720.

Peer, S., Vybornova, A., Saracevic, Z., Krampe, J., Zessner, M., Zoboli, O., 2022. Enhanced statistical evaluation of fluorescence properties to identify dissolved organic matter dynamics during river high-flow events. Sci. Total Environ. 851, 158016. https://doi.org/10.1016/j.scitotenv.2022.158016.

Peer, S., Vybornova, A., Tauber, J., Saracevic, E., Krampe, J., Zessner, M., Zoboli, O., 2023. To analyze or to throw away? On the stability of excitation-emission matrices for different water systems. Chemosphere 333, 138853. https://doi.org/10.1016/j.chemosphere.2023.138853.

Pucher, M., Wünsch, U., Weigelhofer, G., Murphy, K., Hein, T., Graeber, D., 2019. staRdom: versatile software for analyzing spectroscopic data of dissolved organic matter in R. Water 11, 2366. https://doi.org/10.3390/w11112366.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.R-project.org/.

Rodríguez-Vidal, F.J., García-Valverde, M., Ortega-Azabache, B., González-Martínez, A., Bellido-Fernández, A., 2020. Characterization of urban and industrial wastewaters using excitation-emission matrix (EEM) fluorescence: searching for specific fingerprints. J. Environ. Manag. 263, 110396. https://doi.org/10.1016/j.jenvman.2020.110396.

Rodríguez-Vidal, F.J., Ortega-Azabache, B., González-Martínez, A., Bellido-Fernández, A., 2022. Comprehensive characterization of industrial wastewaters using EEM fluorescence, FT-IR and 1H NMR techniques. Sci. Total Environ. 805, 150417. https://doi.org/10.1016/j.scitotenv.2021.150417.

Sgroi, M., Roccaro, P., Korshin, G.V., Vagliasindi, F.G.A., 2017. Monitoring the behavior of emerging contaminants in wastewater-impacted rivers based on the use of fluorescence excitation emission matrixes (EEM). Environ. Sci. Technol. 51, 4306–4316. https://doi.org/10.1021/acs.est.6b05785.

Shen, J., Liu, B., Chai, Y., Liu, C., Cheng, C., Wu, J., 2021. Characterizing fluorescence fingerprints of different types of metal plating wastewater by fluorescence excitation-emission matrix. Environ. Res. 194, 110713. https://doi.org/10.1016/j.envres.2021.110713.

Wang, Y., Hu, Y., Yang, C., Wang, Q., Jiang, D., 2019. Variations of DOM quantity and compositions along WWTPs-river-lake continuum: implications for watershed environmental management. Chemosphere 218, 468–476. https://doi.org/10.1016/j.chemosphere.2018.11.037.

Yan, Y., Li, H., Myrick, M.L., 2000. Fluorescence fingerprint of waters: excitation-emission matrix spectroscopy as a tracking tool. Appl. Spectrosc. 54, 1539–1542. https://doi.org/10.1366/0003702001948475.

Yang, L., Shin, H.-S., Hur, J., 2014. Estimating the concentration and biodegradability of organic matter in 22 wastewater treatment plants using fluorescence excitation emission matrices and parallel factor analysis. Sensors 14, 1771–1786. https://doi.org/10.3390/s140101771.

Yang, L., Han, D.H., Lee, B.-M., Hur, J., 2015. Characterizing treated wastewaters of different industries using clustered fluorescence EEM–PARAFAC and FT-IR spectroscopy: implications for downstream impact and source identification. Chemosphere 127, 222–228. https://doi.org/10.1016/j.chemosphere.2015.02.028.