

Research article

Guidance on effective exploitation of spectroscopic techniques for assessing wastewater discharges in water bodies

Rita Salgado Brito^{a,*}, Elsa Mesquita^a, Rui M.C. Viegas^a, Joana Soares^a,
Cláudio Costa^b, Maria João Rosa^a

^a Urban Water Unit, Hydraulics and Environment Department, National Laboratory for Civil Engineering (LNEC), Av Brasil 101, 1100-066, Lisboa, Portugal

^b TRATAVE, Tratamento de Águas Residuais do Ave S.A., Rua ETAR de Serzedelo, Serzedelo, 4765-543, Portugal

ARTICLE INFO

Keywords:

Early warning
Smart monitoring
UV-Vis spectroscopy
Urban streams
Urban drainage
Wastewater discharge

ABSTRACT

Environmental protection is increasingly critical in urban drainage management. While treated wastewater discharges are routinely monitored, the detection of unexpected or unauthorized drainage inflows into surface waters remains a challenge – especially as discharge points are often undocumented, irregular in time and space, occur under highly variable conditions, and monitoring conditions are far from standard. This is one of the emerging challenges of the 2024 European Directive on Urban Wastewater Treatment. Spectrophotometric techniques, more common in wastewater treatment plants, but less common in drainage systems and water bodies, allow for real-time, in-situ water quality monitoring, even for low-pollution discharges like treated effluents. This paper presents a novel, stepwise methodology for using spectroscopic techniques to locate and characterize urban discharges into water bodies, proposes a new Water Quality Label index, and comprehensively guides their use. Based on UV-Vis absorbance and Fluorescence spectroscopy, the application of visual data mining, differential spectra, isobestic points, PCA, PLS and PARAFAC was demonstrated for this purpose. The methodology was applied in a case study of a treated wastewater discharge, with lower contaminant concentrations than those typically associated with unregistered illicit drainage discharges. By detecting subtle changes in river water composition, spectrophotometry helps to locate unknown discharges and support treatment plant operation. This approach improves the understanding of the interactions between urban drainage dynamics, treated wastewater discharges and receiving water quality and supports monitoring and control efforts and integrated urban water management, in line with the new EU Directive and paving the way to discharge limits based on receiving water quality.

1. Introduction

The protection of surface water from pollutant discharges is increasingly important. On the edge of the 21st century second quarter, the ecological status of Europe's rivers remains a serious environmental concern, impacting biodiversity, public health, and water security. In 2021, only 37% of European surface waters achieved a good or high ecological status, with most impaired by pollution, habitat degradation, and other human-induced pressures (EEA, 2024). Widespread impairment of American surface waters is also reported, with many rivers and lakes failing to meet the Clean Water Act water quality standards (EPA, 2024). Australia and New Zealand provide guidelines for water quality

management and report salinity and land-use as major pressures on surface water quality, (AGI, 2026). Key pressures include agricultural runoff – where excessive use of fertilizers and pesticides leads to nutrient enrichment and chemical contamination; industrial pollution – where chemicals, heavy metals or biochemical contamination may introduce persistent pollutants into water bodies; and urban wastewater – where untreated or inadequately treated wastewater releases increased organic load, microbiological contamination, or contaminants of emerging concern (CECs).

From the urban water utilities' perspective, monitoring the discharges from Wastewater Treatment Plants (WWTPs) is a legal obligation grounded in several EU directives, as WWTPs are critical for

* Corresponding author. Urban Water Unit, Hydraulics and Environment Department, National Laboratory for Civil Engineering (LNEC), Av Brasil 101, Lisboa, Portugal

E-mail addresses: rsbrito@lnec.pt (R.S. Brito), emesquita@lnec.pt (E. Mesquita), rviegas@lnec.pt (R.M.C. Viegas), jfsoares@lnec.pt (J. Soares), claudio.costa@tratave.pt (C. Costa), mjrosa@lnec.pt (M.J. Rosa).

<https://doi.org/10.1016/j.jenvman.2026.129759>

Received 18 February 2026; Received in revised form 15 April 2026; Accepted 18 April 2026

Available online 2 May 2026

0301-4797/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

protecting aquatic ecosystems and public health, by treating urban and industrial wastewater before being discharged into the environment. In addition to these, it is crucial to detect any anomalous inflows into streams – regardless of their origin – to ensure a timely response. Untreated wastewater can reach surface water bodies due to several illegal or unregulated discharges, such as direct sewer connections, from households, including greywater (kitchen, bath) and blackwater (toilet waste), from industries or common commercial facilities (restaurants, gas stations, laundries) or from WWTPs or pumping stations, due to operational constraints. Even stormwater runoff, carrying rainwater washed from streets and urban surfaces, often carries pollutants, such as heavy metals, oil and grease. Combined sewers, designed to carry both sanitary and stormwater flows, are often equipped with overflow devices, which can unintentionally release high pollutant loads during heavy rainfall or when a pipe obstruction occurs downstream.

Regulation of combined sewer overflows (CSO) differs. For example, in the U.S.A., discharges from CSO are regulated through the Clean Water Act, which imposes reporting obligations for each discharge. In contrast, CSO reporting in the EU is less standardized and often depends on national implementation rather than EU-wide legal requirements. In Europe, a great effort will be made to enhance both the understanding and the control of discharges from drainage networks into receiving water bodies, following the revised European Directive on Urban Wastewater Treatment (UWWTD, [Directive 2024/3019](#)) and others aimed at protecting water quality across different contexts, such as ([Directive 2008/105/EC](#)), which sets Environmental Quality Standards (EQS) (see Supplementary Material).

Ultraviolet–visible (UV-Vis) spectroscopy has become an established technique for water quality monitoring and process control, offering rapid, reagent-free operation, minimal sample preparation, and the capability for continuous, real-time measurements ([Thomas and Burgess, 2022](#)). In WWTPs, spectroscopic methods – most notably on-line submersible UV-Vis absorbance sensors – have been widely implemented for process control ([Vanrolleghem and Lee, 2003](#); [Shi et al., 2022](#)), while their use within drainage networks remains comparatively limited. Over nearly two decades, these approaches have demonstrated substantial effectiveness in characterizing wastewater and stormwater quality ([Platikanov et al., 2014](#); [Brito et al., 2016](#); [Chen et al., 2021](#); [Rong et al., 2024](#)), although their performance can be influenced by factors such as sensor fouling, calibration requirements, and variations in matrix composition ([Causse et al., 2017](#); [Guo et al., 2020](#)).

In addition to UV-Vis absorbance spectroscopy, 2D or 3D fluorescence spectroscopy, widely used in biological sciences ([Christensen et al., 2006](#)), is gaining attention in environmental monitoring due to its higher sensitivity and specificity compared with absorbance spectroscopy ([Carstea et al., 2020](#)).

Both UV-Vis absorbance or fluorescence techniques allow for rapid, in-situ or ex-situ monitoring of water quality. Within sewer systems, the spectral signature is highly dependent on the nature of the wastewater ([Pouet et al., 2004](#)). Stormwater inflows typically dilute dissolved substances while increasing suspended solids, leading to detectable changes in the UV-Vis absorbance spectrum. Even though stream water quality can vary considerably, particularly after rainfall events or polluted discharges, UV-Vis absorbance spectroscopy has been applied successfully in these contexts ([Assaad et al., 2017](#); [Zhang et al., 2022](#); [Brito et al., 2022](#); [Ritson et al., 2022](#)).

As such, UV-Vis absorbance spectra (UV-Vis spectra) can serve as valuable indicators of water quality across the urban water cycle. A simple water quality test was proposed ([Thomas and Burgess, 2022](#)) to classify the type of water in natural brooks based on absorbance at three key wavelengths: 210 nm (associated with nitrate presence), 240 nm (used to distinguish dissolved organic matter from suspended solids), and 320 nm (indicative of suspended solids). In a scale from 1 to 4, it codes from natural water with low organic matter or nitrate content, to raw wastewater, respectively.

UV-Vis fluorescence spectra are commonly recorded using excitation–emission matrices (FEEM spectra), generating fluorescence intensity contour maps from emission scans across excitation wavelengths ([Coble, 1996](#)). The spectral data – namely the excitation wavelength (λ_{ex}) required to occur fluorescence and the emission wavelength (λ_{em}) of the emitted signal – allow identification of distinct organic fluorophores. In aquatic environments, there are typically 4 families of fluorescent organic compounds, associated to specific fluorescent peaks – known as Coble peaks ([Coble, 1996](#)) – proposed as indicators of biogeochemical origin and transformation state of dissolved organic matter (DOM): (i) very complex or complex molecules related to soil DOM (humic-like compounds, with Coble peaks C and A_C); (ii) complex molecules that result from microbial products (microbial humic-like compounds, with peaks M and A_M); (iii) small and simple organic compounds that directly derive from microbial activity (tryptophan containing protein-like compounds, with peak T); (iv) and tyrosine containing protein-like compounds (with peak B).

2D-based fluorescence indices provide further insight into DOM composition and source. The humification index (HIX) ([Zsolnay et al., 1999](#)), fluorescence index (FI) ([McKnight et al., 2001](#)) and biological index (BIX) ([Huguet et al., 2009](#)) are the most widely used. HIX indicates the extent of DOM humification or degradation, associated to DOM from soils, vegetation, and upstream runoff ([Zsolnay et al., 1999](#)); FI indicates DOM complexity, with terrestrially or microbially derived sources ([McKnight et al., 2001](#)); and BIX, for biologically derived DOM, associated with microbial activity, algal production, or biologically treated effluent discharges ([Huguet et al., 2009](#); [Jeon et al., 2022](#)).

Parallel factor analysis (PARAFAC) deconvolutes 3D FEEM data into distinct, chemically meaningful components, allowing quantification through normalized score–concentration relationships ([Yang et al., 2015](#); [Soares, 2025](#)). Hence, FEEMs provide a valuable tool for monitoring changes in the composition of dissolved organic matter in water.

Both UV-Vis absorbance and fluorescence spectroscopy techniques show great promise for delivering rapid, straightforward assessments by correlating spectral data with key water quality parameters, including total suspended solids (TSS), chemical oxygen demand (COD), total organic carbon (TOC), dissolved organic carbon (DOC), and biochemical oxygen demand (BOD) ([Brito et al., 2013](#); [Carstea et al., 2016](#); [Zhang et al., 2022](#); [Ritson et al., 2022](#)). Even without quantifying these parameters, the shape of the UV-Vis absorbance spectrum and the FEEM contour plots can signal significant shifts in water quality – due to e.g. organic load variations from stormwater intrusion into sanitary sewers ([Brito et al., 2013](#)) or industrial wastewater inputs ([Pouet et al., 2004](#); [Brito et al., 2017](#); [Carstea et al., 2016](#); [Thomas and Burgess, 2022](#)). In such cases, the spectral profile itself becomes a diagnostic tool, capable of detecting and tracking changes ([Brito et al., 2017](#); [Assaad et al., 2017](#)) or detecting outliers (e.g., through Principal Component Analysis, PCA) ([Booksh, 2000](#); [Brito et al., 2022](#)), or estimating water quality parameters (e.g. through Partial Least Squares regression, PLS) ([Van der Broeke et al., 2006](#); [Torres and Bertrand-Krajewski, 2008](#); [Brito et al., 2013](#); [Viegas et al., 2022](#)). Chemometric methods use advanced multivariate analysis, such as PCA, to enhance and interpret spectral data. PCA identifies patterns and trends by reducing data dimensions, representing spectra through scores of principal components, and highlighting similarities or outliers within a dataset. PLS also reduces data dimension, building predictive models that relate input data (e.g., spectra) to output variables (e.g., water quality parameters).

Despite the powerful individual abilities and complementary potential of spectroscopic techniques, guidance for its integrated and progressive use by water practitioners is still limited. This paper aims to address this gap, by proposing a methodology for applying these techniques, not commonly used by drainage utilities, to locate and characterize discharges from drainage networks or WWTPs into water bodies. A stepwise monitoring strategy of increasing complexity is proposed, building on prior suspicion or knowledge of discharge locations – offering a novel, scalable approach to environmental assessment. Aiming

to provide practical guidance, this paper presents the contextualization and interpretation of results obtained from techniques of increasing complexity – namely visual data mining, differential spectra, isosbestic point analysis, and multivariate methods (PCA, PLS, and PARAFAC – applied to UV-Vis absorbance and FEEM spectra. UV-Vis and fluorescence spectroscopy were selected as these are rapid, non-destructive, and cost-effective techniques, widely used for the characterization of wastewater-impacted waters. In addition, FEEM combined with PARAFAC provides chemically interpretable components, while UV-Vis data can be readily integrated into predictive chemometric models. Although other emerging optical monitoring tools are also promising, such as hyperspectral sensors or in situ probes (Pu et al., 2017; Kumar et al., 2024), these were beyond the scope of the present study.

A case study of treated wastewater discharges in rivers intends to demonstrate the application and limitations of the proposed methodology's ability to improve conventional approaches by enabling rapid detection of water quality changes in the receiving environment, even at trace concentrations of treated effluent. This is key to paving the way for more effective, efficient, and sustainable wastewater treatment in line with the Water Framework Directive (WFD) and the EQS, based on the assimilative capacity of water bodies rather than fixed emission limits.

This approach aims to enhance the understanding of the interactions between urban drainage dynamics, treated wastewater discharges and receiving waters, and to support the implementation of early warning, monitoring, and control measures, ultimately contributing to integrated urban water management, as required by the new UWWTD ((Directive 2024/3019).

2. Materials and methods

2.1. Procedures for spectral data processing and interpretation

The shape of the UV-Vis absorbance spectrum or contour plots of FEEM spectrum can be highly informative on their own.

Preliminary data exploitation can be performed through *visual data mining*, aiming at discovering patterns and relationships and improving understanding of datasets through visualization of graphs, where trends, clusters, and anomalies can be easily spotted (Oliveira and Levkowitz, 2003). For example, when subtracting an UV-Vis spectrum by another, the *differential spectrum* provides relevant information on in-between events (Thomas and Burgess, 2022).

Convergence of all spectra at a single point, known as *isosbestic point* (IP), suggests a mixture of the same two compounds (Pouet et al., 2004). Sometimes, a *hidden isosbestic point* (HIP) might be revealed through *spectral normalization* (Thomas and Burgess, 2022). HIP might be absent when industrial discharges occur in a river (Pouet et al., 2004). *Pre-processing* might be necessary, e.g. mean centering (Coombes et al., 2007).

Statistical and chemometric methods are widely used to extract meaningful information. *Principal Component Analysis* (PCA) examines large data sets, such as spectra, and has been used to detect outliers or highlight data trends (Miller, 2000). It transforms original data into a new set of principal components (PC).

The *Partial Least Squares* (PLS) regression is based on a decomposition method, like PCA, and seeks a reduced number of latent variables (LVs, similar to PCs) that best explain the variability of an X matrix (for example, several spectra) to explain a Y matrix (for example, water quality parameters for each spectrum in the X matrix) (Miller, 2000; Lourenço et al., 2008).

Supplementary Material provides additional information.

2.2. Methodology proposed for detecting or monitoring urban discharges

The proposed methodology is intended for water professionals who suspect an anomalous or uncontrolled discharge in a specific river stretch, as well as for those responsible for monitoring and managing the

impact of known discharges.

To support this, the methodology includes a range of spectral processing and interpretation procedures, adapted to the available knowledge about the discharge's location and characteristics – as illustrated in Fig. 1. As detailed in Supplementary Material, these procedures may involve or not collecting samples at the potential discharge (D), as well as upstream (UP) and downstream (DS) of it, may require sample pre-processing, and range from direct visual data mining, calculation of differential spectra, IPs or HIPs, to statistical analyses.

The proposed procedures range from 1 to 8, with increasing complexity. Procedures 1 to 5 are based on graphical representations of spectra or spectral elementary calculations, whereas procedures 6 to 8 are based on statistical methods. These require access to different information. For instance, PCA, PLS, and PARAFAC are recommended when the discharge is confirmed, but its location is either confirmed or not (PCA). These techniques can be applied when the discharge's water quality parameters are known (PLS, PARAFAC), or when its probable main compounds are likely identified (PCA). In other words, for monitoring planning, the application of PLS and PARAFAC implies that the discharge point is accessible and samples can be collected.

Table 1 summarizes the proposed procedures to apply to the different monitoring objectives and approaches.

As mentioned, *visual data mining* (Fig. 1, procedure 1) can be a first step for a preliminary identification of changes, patterns or clusters in UV-Vis absorbance spectra. Based on certain spectral characteristics (such as peaks, shoulders, or residual absorbance at specific wavelengths) natural water can be classified as originating from urban, mixed, or rural brooks. Urban brooks typically exhibit absorbance values above 0.5 a.u. cm^{-1} below 220 nm, whereas rural brooks generally maintain absorbance below 0.5 a.u. cm^{-1} across the entire UV-Vis wavelength range (Delpla, 2011). Comparison of spectral shapes along a stream enables the location and characterization of an in-between discharge. Visual analysis can be sustained by determination of the natural water's quality label (WQL) proposed herein as an adaptation of the water quality code from Thomas (1995) (cited by Thomas and Burgess, 2022). It relies on UV absorbance measured at three diagnostically relevant wavelengths: 210 nm, associated with nitrite/nitrate presence; 240 nm, used to differentiate DOM from suspended solids; and 320 nm, sensitive to the presence of suspended solids. Variations in pH, temperature, and biological activity, as well as spatial heterogeneity in DOM sources and environmental conditions, may influence absorbance measurements (Jaffé et al., 2008; Erlandsson et al., 2012; Ma and Li, 2020).

Despite its simplicity, this approach provides valuable preliminary insights into the origin and composition of the water source, by integrating (i) the absorbance intensity values at three selected wavelengths and (ii) the typical relative contributions of DOM, inorganic constituents (mainly nitrate, nitrite, and ferric or chromium complexes), and suspended materials (such as organic colloids and fine mineral particles) to the overall water's UV-Vis absorbance (Thomas and Burgess, 2022, as expressed in Table 2). When different codes may be assigned in the same spectrum, herein it is proposed to attribute the lower of those codes to each absorbance value and subsequently calculate a natural *water quality label* (WQL) by Eq. (1), for each spectrum:

$$WQL = \frac{\sum C_i}{12} \quad (\text{Eq. 1})$$

where:

WQL: water quality label (–)

Ci: assigned water quality code, from 1 to 4, considering the absorbance values ($A_{\lambda,i}$) at wavelengths λ of 210 nm, 240 nm and 320 nm, according to Table 2.

WQL can vary between 0.25 and 1.00 (the higher, the lower the water quality) and the following interpretation is proposed: (i) between 0.25 and 0.375, the absorbance spectrum refers to a natural water with

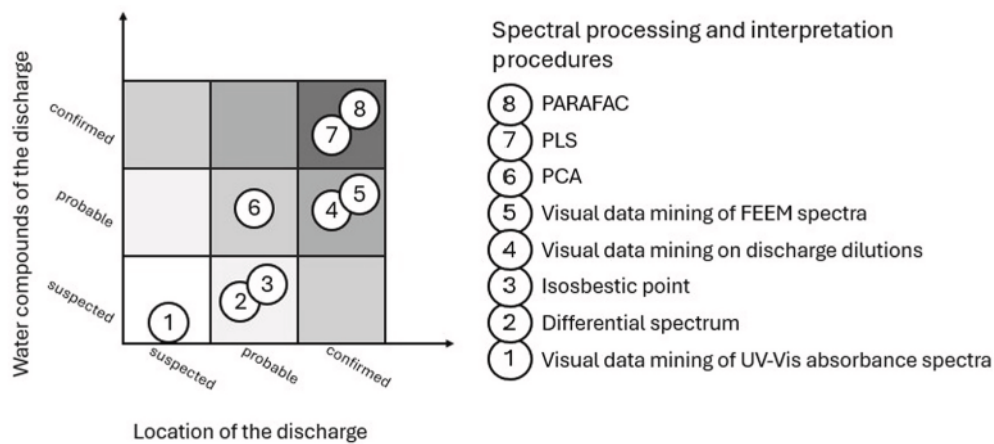


Fig. 1. Recommended spectral processing and interpretation procedures depending on increasing knowledge regarding the location of a discharge into a river and its water quality: 1) Visual data mining of UV-Vis absorbance spectra; 2) Differential UV-Vis spectrum; 3) Isosbestic point (or hidden isosbestic point, if required) in UV-Vis absorbance spectra; 4) Visual data mining on discharge dilutions, based on UV-Vis spectra; 5) Visual data mining of FEEM spectra; 6) PCA of UV-Vis absorbance spectra; 7) PLS of UV-Vis absorbance spectra; 8) PARAFAC of FEEM spectra.

no (or with very low concentration of) DOM and nitrate or to treated and denitrified wastewater; (ii) between 0.375 and 0.625, to a natural water with nitrate or highly treated biological WWTP outlet; (iii) between 0.625 and 0.875, to a water similar to regular biological WWTP outlet; (iv) between 0.875 and 1.00, to a water similar to untreated wastewater. Nevertheless, given the sensitivity of absorbance to environmental variability, these WQL must be used with caution, as a preliminary indicator.

Visual analysis can also be supported by the *differential spectrum* (Fig. 1, procedure 2), obtained by absorbance subtraction for the same wavelengths, between upstream and downstream samples. The peaks and lumps hinted by visual data mining can be isolated and related to specific spectral changes associated with substances added in between.

As mentioned in Table 1, preliminary spectral *normalization* may be required for some of the procedures, which consists of assigning to all spectra the same underlying area, according to Eq. (2):

$$A_i^* = \frac{A_i \cdot N}{\sum (A_i \cdot \Delta\lambda)} \quad (\text{Eq. 2})$$

where:

A_i^* : corrected absorbance for wavelength i (a.u.)

A_i : measured absorbance for wavelength i (a.u.)

N : area under the spectrum to be imposed by normalization (e.g., 100 nm)

$\Delta\lambda$: wavelength step (e.g., 1 nm).

Isosbestic points (IP) in UV-Vis spectra (Fig. 1, procedure 3) refer to a specific wavelength or region at which the absorbance of several water samples remains constant, i.e., where all spectra cross in the graph. The presence of an IP (or hidden isosbestic point, HIP) indicates that changes between the samples involve only two principal components that are interconverting. When applied to upstream and downstream spectra, the presence of an IP or HIP indicates that no relevant substance (different from the others already present) was added in-between. These can be detected visually in the graph, through determination of the *differential spectrum* (Fig. 1, procedure 2) and the identification of a single location with residual absorbances close to zero, and can be simply assessed through the calculation of the *spectral variability* (Thomas and Burgess, 2022), as in Eq. (3). The lower the spectral variability V , the stronger the evidence for an IP or HIP.

$$V = \left(1 - \frac{N_{IP}}{N_t}\right) \cdot 100 \quad (\text{Eq. 3})$$

where:

V : spectral variability regarding an IP or HIP (%)

N_{IP} : number of spectra crossing the IP or HIP (—)

N_t : total number of spectra (—)

To estimate the *dilution* of a given discharge (Fig. 1, procedure 4), visual data mining can be used to test a hypothesis. The real downstream UV-Vis spectra (DS) can be compared to several spectra acquired in different composite samples for varying dilutions (Dil_i) of the discharge effluent (D) on the upstream river (the several hypotheses under test). DS spectrum will be positioned between the spectrum of Dil_α and Dil_β , meaning that the river downstream (DS) is composed of approximately $\alpha\%$ to $\beta\%$ of discharge effluent (D), diluted in the upstream river water (UP).

Visual data mining of UP and DS fluorescence spectra enable detecting relevant discharges of DOM (Fig. 1, procedure 5). 2D excitation or emission spectra, 3D excitation–emission matrix (FEEM) contour plots — fluorescence intensity as a function of excitation (λ_{ex}) and emission (λ_{em}) wavelengths — or peak analysis can be used. Fluorescence emission signals at longer excitation and emission wavelengths link to greater molecular complexity, whereas shorter wavelengths correspond to simpler organic compounds (Zsolnay et al., 1999). Table 3 resumes fluorescence peaks of DOM commonly found in aquatic environments, their characteristics and possible sources.

Fluorescence indices, calculated as ratios of fluorescence intensities at selected emission wavelengths or wavelength ranges under specific excitation wavelengths, offer complementary information and, when jointly interpreted, provides information on DOM composition and potential sources. The most widely used (Table 4) distinguish between microbially derived material and terrestrially derived inputs, and assess molecular complexity, aromaticity, and DOM humification degree. Presented values are indicative as may vary with environmental conditions (Serène et al., 2025). If the aquatic environment differs from that used to develop the index, its excitation and emission wavelengths may need minor adjustments (Serène et al., 2025).

As stated above, *Principal Component Analysis* (PCA, Fig. 1, procedure 6) transforms the original data (UV-Vis absorbance spectra) into a new set of variables (principal components, PC1, PC2, ...). For PCA, spectra were preprocessed and outliers were identified using score plots (showing the first two PCs) as well as Hotelling's T^2 and Q-residuals plots (details on the procedures in Supplementary Material). Samples falling outside the 95% confidence interval in any of these plots may be flagged as outliers. These are then examined to determine whether they correspond to an underrepresented normal sample (in which case it is kept in the model) or if other evidence exists to distrust the sample, e.g. in campaigns registry or in water quality parameters (in which case it is

Table 1
Methodology proposed for monitoring, processing, and interpreting spectroscopic data according to the monitoring objective.

PROCEDURE ID.		
1 VISUAL DATA MINING OF UV-VIS SPECTRA (absorbance)		
Objective: To locate a relevant discharge and screen its major components		
Monitoring approach	Data processing · Might need to mean center absorbance spectra to account for natural river dilution. · Plot on the same graph the absorbance spectra, UP and DS a given location. · Visually identify any changes in spectra shape.	Data interpretation · Changes in spectral shape indicate the introduction of a different substance at that location; · In such cases: calculate and plot WQL (Eq. (1)) to assess the significance of absorbance changes.
2 DIFFERENTIAL UV-VIS SPECTRUM		
Objective: To profile the major components of the discharge		
Monitoring approach	Data processing · Acquire UV-Vis spectra UP and DS of the potential discharge	Data interpretation · Relate relevant shoulders in Diff with corresponding parameters in that spectral window.
3 UV-VIS ISOBESTIC POINTS		
Objective: To locate a relevant discharge and screen its major components		
Monitoring approach	Data processing · Acquire UV-Vis spectra along the river	Data interpretation · HIP occurs: any discharge in between has the same major compounds as the river. · HIP doesn't occur: different substance was added in between, in that location. Proceed to a more robust procedure.
4 VISUAL DATA MINING ON DISCHARGE DILUTIONS SPECTRUM		
Objective: To assess the discharge dilution by river stream		
Monitoring approach	Data processing · Collect UP, DS and D water · Prepare Dill: D serial dilutions with UP water. · Acquire UV-Vis spectra of DS, D and Dill	Data interpretation · Dilution values of Dill spectra adjacent to the DS spectrum provide an estimate of the river's discharge dilution.
5 VISUAL DATA MINING OF FEEM SPECTRA		
Objective: To locate a relevant discharge and improve screening of its major components		
Monitoring approach	Data processing · Collect UP, DS and D water · Filter and dilute with ultrapure water, if necessary. · Measure FEEM spectra.	Data interpretation · Changes in FEEM indicate the introduction of a different DOM content at that location; · In such cases: calculate fluorescence indices (FI, BIX, HIX).
6 PCA OF UV-VIS SPECTRA		
Objective: To locate a relevant discharge and trace its likely source		
Monitoring approach	Data processing · Acquire UV-Vis spectra upstream (UP) and downstream (DS) of the potential discharge point.	Data interpretation · If separate clusters for UP and DS occur: a different substance was added in between. · If DS is close to a known cluster, that's likely the source.
7 PLS OF UV-VIS SPECTRA		
Objective: To estimate water quality parameters of the river		
Monitoring approach	Data processing · Acquire UV-Vis spectra at UP and DS of the potential discharge point · Analyze regular water quality parameters	Data interpretation · If model's results are acceptable: estimate of water quality parameters based on UV-Vis data.
8 PARAFAC OF FEEM SPECTRA		
Objective: To evidence the discharge's existence and impact, and identify its possible source		
Monitoring approach	Data processing · Examine UP, DS and D FEEM contour plots to identify raw variations and shared fluorescence regions. · Filter and dilute with ultrapure water, if necessary. · Measure FEEM spectra.	Data interpretation · UP vs DS FEEM differences indicate a new DOM component between points. · Shared D and DS fluorescence regions, absent in UP, suggest D impacted river water quality.

Table 2

Natural waters' quality code based on absorbance at 210 nm (A_{210}), 240 nm (A_{240}), and 320 nm (A_{320}) and rationale behind code's interpretation (adapted from Thomas and Burgess, 2022).

Absorbance (a.u.cm ⁻¹)			Code	Interpretation
A_{210}	A_{240}	A_{320}		
<0.5	<0.2	<0.05	1	Natural water with no (or very low) DOM and nitrate; treated, denitrified wastewater
0.5 - 1.0	<0.2	<0.05	2	Natural water with nitrate or efficient biological WWTP outlet
0.5 - 1.0	0.2 - 0.5	>0.2	3	Activated sludge-WWTP outlet
>1.0	>0.5	>0.2	4	Untreated wastewater

Rationale.

A_{210} – similar contribution of dissolved inorganics (mainly nitrates/nitrites) and organic matter (mainly low molecular organics, proteins and conjugated carbonyl groups), little contribution of suspended material (except through scattering).

A_{240} – dominated by DOM, mainly aromatic compounds or with conjugated bonds; low contribution of dissolved inorganics or suspended material (except in high turbidity waters).

A_{320} – dominated by DOM, mainly dissolved humic and fulvic acids; suspended organic matter can also absorb and scatter light; low contribution of inorganics (only certain dissolved ferric or chromium complexes adsorb, with minor scattering from suspended inorganics).

removed). When representing each spectrum in a scores plot (a PC1/PC2 graph), similar spectra will cluster. If upstream and downstream spectra cluster jointly, it means they are essentially similar. If they cluster separately, then a relevantly different substance was added in-between. When spectra of a probable wastewater source are available (e.g. an upstream industry), including them in the PCA model allows to check clustering with downstream spectra, evidencing it as a likely discharge (Brito et al., 2017).

Models are typically built using a training set (two-thirds of available data) and then validated with an external test set (remaining one-third). However, when working with smaller datasets, all available data may be used for training, and cross-validation estimates the model's predictive performance. In this study, since each campaign included fewer than 15 spectra per section, the full dataset was used as training set. It was evaluated using leave-one-out cross-validation, which is advisable when the dataset is very small and no independent test set is available. As there is a risk of overfitting, the resulting model is considered a preliminary validation, not a definitive proof of model performance. To reduce overfitting risk, it is important to limit model complexity and use fewer principal components.

The root mean squared error (RMSE, Eq. (5)) of cross-validation was determined. The optimal number of PCs was the one that captured data variance above 99%, while minimizing RMSE.

$$RMSE = \sqrt{\frac{\sum (O_i - P_i)^2}{n}} \quad (\text{Eq. 5})$$

where:

n: number of samples.

O_i : observed water quality parameter for sample i.

P_i : predicted water quality parameter for sample i.

As for *Partial Least Squares* (PLS, Fig. 1, procedure 7), it can be used as a multivariate statistical method to model the relationship between two datasets, X and Y (e.g. UV-Vis spectra and water quality parameters like TOC, COD, etc.). This predictive model links changes in the spectra to water quality parameters. With a robust PLS model, a rapid estimate of river water parameters (Y) can be made based only on spectra (X). PLS was used to predict COD, NO_3 , TOC and conductivity values from river UV-Vis data, which was preprocessed using the Savitzky-Golay filter and mean centering. A leave-one-out cross-validation procedure was applied with the same conditions as those referred to the PCA models and, as to reduce overfitting, the number of latent variables was kept to the minimum required to minimize RMSE. The optimal number of LVs was the one minimizing the RMSE. Model adequacy was determined by calculating relative RMSE (RMSE_{rel} , in %) for each water quality parameter, i. e., RMSE related to the average observed values of the parameter (Eq. (6)).

$$\text{RMSE}_{\text{rel}} = \frac{\sqrt{\frac{\sum (O_i - P_i)^2}{n}}}{\bar{O}} \cdot 100 \quad (\text{Eq. 6})$$

where:

n: number of samples.

O_i : observed water quality parameter for sample i.

P_i : predicted water quality parameter for sample i.

$\bar{O}$: average observed values for all samples

PARAFAC (Fig. 1, procedure 8) can be applied to decompose FEEM data and identify the fluorophores in the water. A key advantage of PARAFAC over other statistical methods is that it preserves the three-dimensional structure of the dataset, allowing identification of chemical components with physically interpretable meaning. Before performing PARAFAC analysis, exploratory data assessment must be conducted. Proper selection of emission and excitation wavelength ranges will reduce noise and enhance overall data quality. Sample outliers should be identified and removed from data set, using Hotelling's T^2 and Q-residuals plots (Stedmon and Bro, 2008). Several data processing procedures must be applied during model fitting, namely non-negativity constraint, Random orthogonal initialization and sample shuffling. The validation procedures of PARAFAC model include split-half analysis and core consistency diagnostics (Murphy et al., 2013; Stedmon and Bro, 2008). In this study, PARAFAC models with increasing numbers of components were tested, and the final model was selected based on split-half validation, core consistency diagnostics, and the chemical interpretability of the excitation and emission loadings.

Simple linear regressions were established between PARAFAC component scores and measured water quality parameters using the full dataset available for each sampling set. Because of the limited number of samples, no independent calibration/test split was applied. Model performance was therefore evaluated using the coefficient of determination (R^2) and RMSE_{rel} .

Successful implementation of the proposed methodology relies on appropriate monitoring campaigns, reliable water analysis results, and adequate data analysis methods and tools.

2.3. Monitoring stations and sampling campaigns

The approach was tested in Lordelo case study, located in Vizela River, northern Portugal, near Lordelo WWTP discharge, within LIFE Fitting project.

The Vizela river is regulated by several dams, some integrated with pumping stations, which significantly influences the hydraulic behavior of the natural water system. The Lordelo wastewater system comprises the sewer network and a WWTP – daily treatment capacity of approximately 28,000 m³, both industrial and urban wastewater. Lordelo WWTP scheme includes pre-treatment (pumping, screening, grease and

Table 3

Common fluorescence peaks from DOM in aquatic environments (Stedmon et al., 2003; Coble, 2007; Serène et al., 2025).

Description	$\lambda_{\text{ex,max}}$	$\lambda_{\text{em,max}}$	Peak	Common DOM source
Pigment like	398	660	P	Very complex DOM
Soil-humic-like	455	521	E	Very complex DOM soil derived
	390	509	D	
	250	504	-	
	320–360	420–460	C	
	260	400–460	A _C	Complex DOM soil derived
Microbial humic-like	290–310	370–410	M	Complex N-rich DOM resulting from aggregation of microbial products
	240	350–400	A _M	
	280	370	N	N-rich DOM resulting from phytoplankton productivity
Tryptophan-like, protein like	275	340	T	Small N-rich DOM directly derived from microbial activity
	230	340	A _T	
Tyrosine-like, protein-like	275	305	B	
	230	305	A _B	

grit removal), secondary treatment (biological treatment via activated sludge), and tertiary treatment for color removal using sand filtration and ozonation.

The drainage system is partially combined, where some stormwater enters the sewer system and reaches the WWTP. Stormwater flow can rapidly change, altering the water quality matrix, compromising the system's capacity, affecting treatment efficiency, and increasing discharges of uncertain quality. Climatically, the region experiences hot, moderately humid summers (June–September) and long, cold, rainy autumns and winters (September–March), with 1500 mm average annual precipitation. Often, Vizela river becomes almost dry, in the Summer. Moreover, industrial contribution decreases in August (some industries close for holidays).

WWTP inflow thus fluctuates diurnally, seasonally, and annually, driven by urban and industrial water uses and precipitation dynamics.

Upstream, some textile companies and other plants are authorized to discharge treated industrial wastewater.

To test whether the proposed procedures could detect a low-impact discharge, the treated effluent of Lordelo WWTP was the selected discharge into the river (D). Figs. 2–3 illustrate the sampling stations' context: WWTP discharge (D) and Vizela river upstream (UP) and downstream (DS) of D. River width varies from ca. 10 m to more than 45 m, depending on the flow.

Nine river sampling campaigns were conducted in Lordelo (Table 5). The first three (C1, C2, C3) occurred in dry weather, while the rest were conducted under wet weather, either immediately after or during rain events. During C2, industrial activity was halted due to summer vacation, leading to changes in WWTP's discharge.

Depending on safety conditions and river accessibility, the number of samples per station varied, from one or two to as many as seven per cross-section, which enabled a better understanding of cross-section

variations. In the WWTP discharge, one sample was collected in each campaign. A total of 21, 20 and 8 samples were made available for UP, DS and D, respectively.

The spectra of these samples were acquired in the laboratory, using a bench spectrophotometer – ex-situ spectra. Additionally, during campaigns C6 to C9, a portable UV-Vis probe allowed in-situ measurements. For each river campaign, 5 to 14 spectra were acquired with the portable probe, totaling 35 UP and 31 DS in-situ spectra (Table 5).

Unstable river eddies exhibiting circular motion were observed. Turbulent flow enhances radial mixing and, consequently, improves sample representativeness. A 1-L flask attached to a pole allowing sampling closer to the centerline, further from the riverbanks – where water velocity was lower. Samples were refrigerated for transport; UV-Vis spectra and quality parameters were analyzed within 24 h.

The proposed procedures were initially applied to the ex-situ spectra. Given that in-situ spectra were only available for latest campaigns, they were used to validate the conclusions' consistency.

2.4. Analytical methods

TOC, nitrate and conductivity analysis followed standard methods (Rice et al., 2012). Ex-situ UV-Vis spectra were acquired with a spectrophotometer (Evolution 201, Thermo Scientific) over the 200–750 nm range at 1 nm intervals, using a 10 mm path length quartz cell. In-situ UV-Vis spectra were acquired using a portable submersible probe (Spectrolyser, Scan) with a 100 mm optical path length for 200–750 nm at 2.5 nm intervals. Whenever needed, water samples were diluted (1:10) with distilled water to overcome signal saturation in the UV region.

Table 4

Fluorescence indices calculation, meaning and interpretation (Zsolnay et al., 1999; McKnight et al., 2001; Huguet et al., 2009; Serène et al., 2025).

Index	Meaning	Values	DOM main characteristics
HIX $\frac{\sum_{\lambda_{\text{em}}=435}^{480} I_{\text{em}}}{\sum_{\lambda_{\text{em}}=300}^{345} I_{\text{em}}}$ at $\lambda_{\text{ex}} = 254$ nm	Complexity of organic matter considering the ratio of very complex molecules (numerous aromatic nuclei, high MW organic matter, humic-like) to simpler molecules (few aromatic nuclei, lower MW, protein-like).	<4 4 – 6 6 – 10 >16	High content in simple DOM Low content in very complex humic-like DOM and important content in simpler DOM Important content in very complex humic-like DOM and low content in simpler DOM High content in very complex humic-like DOM
FI $\frac{I_{450}}{I_{500}}$ at $\lambda_{\text{ex}} = 370$ nm	Indicator of the humic-like C compounds precursors; ratio of very complex to simpler humic-like C compounds	1.3 – 1.5 1.7 – 1.9	Terrestrially derived DOM Microbially derived DOM
BIX $\frac{I_{380}}{I_{430}}$ at $\lambda_{\text{ex}} = 310$ nm	Indicator of humic-like origin: microbially (humic-like M) or lignin/soil derived (humic-like C); can be calculated only if peak M exists	<0.7 0.7 – 0.8 0.8 – 1 >1	Low content in microbially derived DOM Medium content in microbially derived DOM High content in microbially derived DOM Biologically or microbially derived DOM

3. Results and discussion

3.1. Procedure 1 | visual data mining of UV-Vis absorbance spectra

Direct comparison of UV-Vis absorbance spectra at D, UP, and DS (Fig. 5) reveals that most samples share a characteristic shape, with peak absorbance above 0.7 a.u.cm^{-1} below 220 nm and a slight shoulder near 280 nm. This means that at this location the river is a typical urban brook (Delpa, 2011), either UP and DS. While the overall shape is similar, DS spectra show a shift, from the UP spectra towards the D spectra. This distinction between UP and DS is less apparent when all sampling campaigns are shown together (Fig. 4), where DS spectra nearly overlapped UP's. However, the difference becomes clearer when spectra from individual campaigns are shown separately (Fig. 5).

WQL was determined and classified for all spectra (Fig. 6). The results suggest that UP already had a relevant content of nitrate and DOM, resembling the effluent of an activated sludge WWTP – a similarity even more pronounced in DS, as expected.

3.2. Procedure 2 | differential UV-Vis absorbance spectrum

DS and UP differential spectra (Fig. 7, except for C7 – data unavailable) confirm the UP-DS changes suggested by visual data mining.

The differential spectra approach enables to identify changes between samples. As wastewater constituents exhibit overlapping spectral fingerprint regions, not a single sharp peak, interpretation of differential spectra requires an understanding of how specific wavelength ranges relate to distinct wastewater constituents as well as the underlying physicochemical properties governing their absorbance spectrum. To aid interpretation of differential spectra, Table 6 lists typical contributors to UV-Vis absorbance in river water and WWTP effluents, along with common features in differential spectra.

In C1 and C3 (Fig. 9), a distinct peak near 230 nm was observed, likely due to reduced river flow under dry weather, which enhanced the influence of the WWTP discharge. C2, also in dry conditions, but during summer vacation with suspended (or strongly decreased) industrial activity, showed only a shoulder – rather than a peak – at 230 nm. This difference suggests the peak may be associated with a persistent industrial contaminant, though further data is needed for confirmation. Campaigns C3 to C9 occurred during wet weather, when higher river flow and dilution reduced the discharge's impact.

A minor shoulder at 280–290 nm was observed in all campaigns, linked to effluent's organic matter. A 300–340 nm bump appeared only in C9, suggesting distinct discharge behavior in this campaign.

3.3. Procedure 3 | IP and HIP in UV-Vis absorbance spectra

Since no IP were observed, all spectra were normalized to search for HIPs. Joint analysis of normalized UP, DS, and D spectra (Fig. 8, top) shows that D lacks a common HIP, indicating that the WWTP effluent

and river samples involve more than two dominant overlapping compounds.

Conversely, normalized UP and DS spectra (Fig. 8, bottom) reveal a joint HIP, suggesting that despite some differences, the river largely maintains the same primary components longitudinally. The discharge D, though distinct, does not significantly alter the river's fundamental water quality matrix. A spectral variability (V, Eq. (3)) of 6% reinforces this HIP.

Other potential HIPs were checked at higher wavelengths, but none were detected.

Additional HIP investigations were made on complementary spectra subsets: i) UP across all campaigns; ii) DS across all campaigns; iii) UP and DS within each campaign. HIPs emerged in all cases. Spectral variability (V, Eq. (3)) was 11% for (i) and (ii), confirming HIP presence; V was not determined for (iii) due to fewer spectra. The results indicate that, aside from dilution, the river's main water quality components remain relatively stable upstream (i) and downstream (ii) across seasons and for given moments in time (iii), despite variations in WWTP discharge and incoming wastewaters.

3.4. Procedure 4 | visual data mining on discharge dilutions

For each campaign in Vizela river, spectra were acquired in composite samples containing 5, 10, 25, 50, 75 and 90% WWTP discharge (D) mixed with UP. These spectra (Dil5%, Dil10%, etc.) were plotted alongside UP (Dil0%) and D (Dil100%) in Fig. 9 – in shades of blue, from lighter (Dil0%) to darker (Dil100%). DS was also plotted (in red), to estimate the real dilution of the WWTP discharge. DS positioning over Dil25% for C3 (Fig. 9, left), and a bit above Dil10% for C5 (Fig. 9, right), indicate WWTP dilutions of 25% and a bit above 10%, respectively. In the other campaigns, estimated dilutions were ~10% in C1, 5–10% in C2, and below 5%.

3.5. Procedure 5 | visual data mining of FEEM spectra

FEEM contour plots of UP, D, DS samples (Fig. 10) reveal campaign-specific patterns between UP and DS samples and highlight common fluorescence regions in D and DS samples, suggesting that D affected river's DOM, given complex nitrogen-rich organic compounds (microbial humic-like) and simple organic compounds (tryptophan- or protein-like) associated with microbial activity.

Fluorescence indices were calculated to assess fluorescent DOM's source in D, UP and DS, as described in Table 4, being presented in Fig. 11.

HIX values (Fig. 11, left) indicate predominantly non-complex DOM in all samples, remaining relatively stable in UP and DS river samples across campaigns (1.4–1.5), while varying more widely (1.1–2.5) in D. However, this index did not reveal any clear influence of D on receiving waters. FI values (Fig. 11, centre) indicates that complex humic-like DOM was terrestrially derived, even in D. The values' shift from UP to



Fig. 2. River sections upstream (UP, left) and downstream (DS, center) of WWTP discharge (D, right).



Fig. 3. Location of sampling stations: WWTP discharge (D), river upstream (UP) and downstream (DS).

Table 5
Sampling campaigns.

Campaign id.	Weather	Industrial activity	Ex-situ spectra	In-situ spectra
C1 2023.07.06, afternoon	Dry	Yes	2 UP + 2 DS + 1 D	–
C2 2023.08.30, midday	Dry	No	2 UP + 2 DS + 1 D	–
C3 2023.10.12, midday	Dry	Yes	2 UP + 4 DS + 1 D	–
C4 2024.01.04, midday	Wet	Yes	3 UP + 3 DS + 1 D	–
C5 2024.07.11, midday	Wet	Yes	7 UP + 5 DS + 1 D	–
C6 2024.10.03, morning	Wet	Yes	3 UP + 2 DS + 1 D	11 UP + 11 DS
C7 2024.10.14, afternoon	Wet	Yes	–	14 UP + 10 DS
C8 2024.10.15, morning	Wet	Yes	1 UP + 1 DS + 1 D	5 UP + 5 DS
C9 2024.10.15, midday	Wet	Yes	1 UP + 1 DS + 1 D	5 UP + 5 DS

DS toward those of D suggests that humic-like DOM from D affected the receiving water. BIX (Fig. 11, right) UP values were consistently below 0.8, indicating low to intermediate biological or microbial derived DOM. DS showed higher BIX, especially in C5, pointing to stronger microbial DOM component likely influenced by D. D showed the highest BIX, confirming that the treated effluent significantly discharges biologically derived DOM. Overall, this pattern highlights the association between WWTP discharge and the downstream enrichment on N-rich, low-molecular-weight DOM.

Source and fate of compounds disclosed by these indices are indicative and should be interpreted cautiously, as they depend on decomposers' access to organic matter (Lehmann and Kleber, 2015) and on environmental conditions (pH, living conditions of those organisms), affecting fluorescence intensity and peak positions (Serène et al., 2025).

3.6. Procedure 6 | PCA of UV-Vis spectra

PCA models were developed for all campaigns using two approaches to site grouping: (a) UP, DS, D; (b) UP, DS. To prevent bias, even when more spectra per day were available, only two were used for each river station (UP, DS), and for D (Table 4). Limited dry-weather data constrained comparisons with wet conditions. Spectra were mean-centered and Savitzky-Golay preprocessed. Table 7 shows model details.

Few PCs were enough to build a model with high explained variance and low RMSE, indicating robust model performance. The identified

outliers were retained, as there were no other signs of samples' distrust. Fig. 12 presents PC1-PC2 scores for each spectrum.

The scores plot in model a), Fig. 11 (left), shows discharge (D) forming a distinct cluster, clearly separated from river spectra. In contrast, UP and DS river data show no clustering in either model, indicating similar water quality at both stations; a more polluted discharge would be needed to separate them. The similarity of UP waters to WWTP effluent (Fig. 6) may be relevant.

Alternative PCA approaches, like limiting wavelength ranges or modeling a single site with others for validation, gave conclusions consistent with those above.

3.7. Procedure 7 | PLS of UV-Vis spectra

PLS models were developed for UP and DS spectra (X matrix), and for TOC, conductivity, COD and nitrate (Y matrix). Table 8 shows separate models for each parameter. TOC and conductivity performed well, with high Y-variance and RMSE_{rel} below 20%, allowing reasonable parameters' estimates, with a tolerable error. Tests with smaller wavelength windows still gave RMSE_{rel} above 12%, indicating more data are needed for robust PLS models; IPLS could help identify optimal spectral windows. COD and nitrate showed poor correlations, with many LV, low Y-variance, and high RMSE_{rel}, likely due to a less representative parameter range than TOC and conductivity.

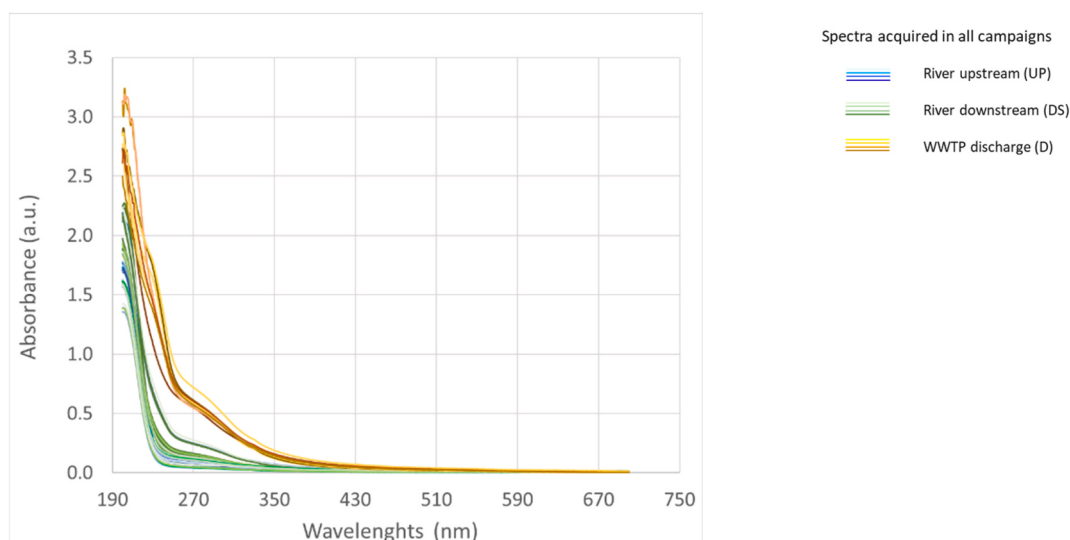


Fig. 4. UV-Vis spectra at D, UP, and DS.

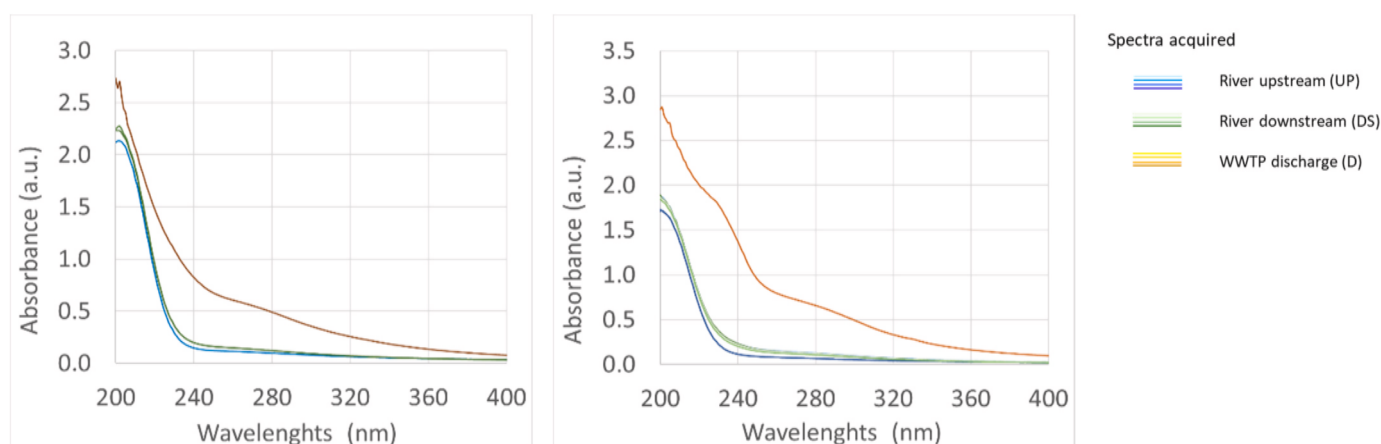


Fig. 5. Examples of UV-Vis spectra, C2 (left) and C5 (right) at D, UP and DS.

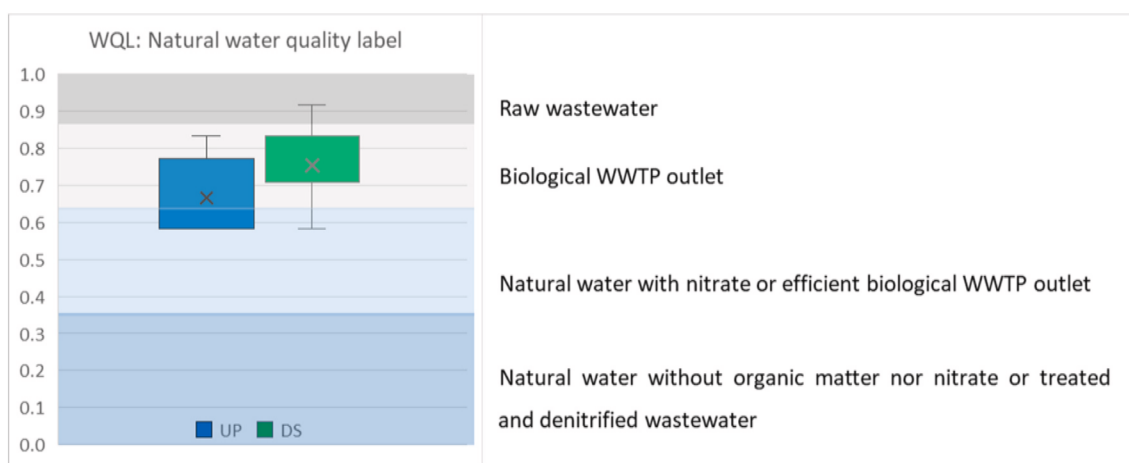


Fig. 6. Box-and-whiskers plot of WQL for all UP and DS spectra.

3.8. Procedure 8 | PARAFAC of FEEM spectra

PARAFAC analysis of UP, D and DS FEEM spectra data identified five components (Fig. 13) matching Table 3 peaks: C and A_C (very complex and complex soil humic-like DOM), M and A_M (microbial humic-like

DOM) and T (Tryptophan-like, protein-like DOM). Each component's relative amount was assessed in all samples from campaigns C1-C6 (except C4), through component scores (Fig. 14). D samples showed the highest scores for all components except peak A_C , which was higher at UP and DS in all campaigns (except C5). The microbial humic-like

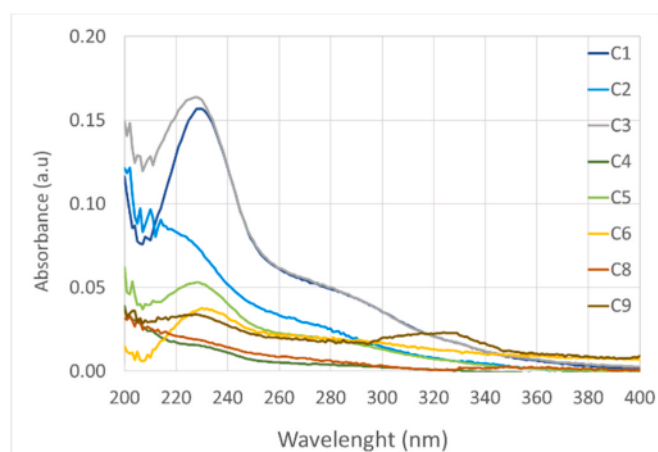


Fig. 7. Differential spectra for each campaign Ci.

were the predominant components in D samples, with peak A_M (representing less complex microbial humic-like DOM) presenting the highest scores under conditions of industrial activity (mostly textile). This likely reflects the chemical transformation of complex organic pollutants (e.g., dyes, surfactants), through ozonization, and further microbial degradation in the WWTP. These oxidative processes generate less complex DOM at the treated effluent, leading to higher fluorescence intensities of peak A_M .

In the river samples, the 5 components' scores shifted from UP to DS, trending toward the values observed at D, likewise the trend of the fluorescence indices FI and BIX.

Component scores from the river, both UP and DS, showed significant correlations with several parameters (Table 9), demonstrating the utility of component scores as surrogates for chemical and microbiological indicators. Particularly, the microbial humic-like (peak A_M) and the soil humic-like (peak D) components showed good correlations, though not always the best, with all parameters except *E. coli*. Table 9 shows the best correlations for each parameter in separate models.

Peak A_M exhibited strong predictive power for TOC and total nitrogen – key indicators of organic pollution and nutrient enrichment – suggesting this component effectively captures the signal of biologically-processed, microbially-derived DOM. A_M also performed excellently in hydrochlorothiazide concentration prediction, highlighting its potential as a non-target pharmaceutical-residues' indicator.

Tryptophan-like fluorescence (peak T) was the best predictor for COD, suggesting that protein-like DOM, linked to microbial activity and effluents, accounts for a significant share of oxygen-demanding organic fraction.

Peak C was the most strongly correlated with total phosphorus, likely reflecting that phosphorus is carried with soil-derived humic substances,

especially during runoff or erosion.

Electrical conductivity was best explained by the microbial humic-like component (peak M), linked to the WWTP effluent samples – with higher conductivity.

Finally, *E. coli* was moderately explained by humic-like component (peak A_C), suggesting that microbial contamination events may be associated with DOM exhibiting classic humic-like spectral features. However, lower variance suggests this is less clear than for chemical parameters.

Although $RMSE_{rel}$ was rather low, an exploratory comparison with another location showed that some fluorescence components were also significantly related to conventional parameters and selected CECs (see Supplementary Material). These results suggest the potential applicability of the approach beyond the studied site, but site-specific validation under different environmental and hydrological conditions is necessary.

4. Conclusions

This paper presents a novel methodology for the stepwise characterization of urban wastewater discharges into water bodies, using spectroscopic techniques, and a novel water quality label. The novelty lies in the stepwise methodology itself. The individual techniques themselves are not new and have been applied in other contexts; it is the systematic integration of multiple methods, rather than any single technique, that provides added value. This combined framework can support utilities in effectively identifying and characterizing polluted discharges, allowing them to allocate appropriate resources to address the specific problems at hand – no more, and no less.

As a step-by-step framework of techniques with increasing complexity, it can support utilities at different stages: those beginning their efforts to protect river water quality, smaller utilities with limited technical capacity, and more advanced users of spectroscopic methods.

Besides being accessible to diverse users, the proposed approach handles complexity progressively (even advanced practitioners can start simple and add sophistication only when needed), improves interpretability of the results, and adapts to data quality, as different steps can be used depending on data availability and reliability.

The approach is applicable in various scenarios, regardless of prior knowledge of discharge's location, accessibility, or water quality. Naturally, when the water quality of the river is unknown or changes rapidly, daily or seasonally, it might be necessary to implement a preliminary campaign to inform the adequacy of the equipment's range (e.g., the optical window of the spectrophotometer), sampling frequency and timing (e.g., to capture storm events or intermittent discharges), or specific sampling procedures (e.g. regarding collected water volume, sample dilution).

The procedures were tested in a case study involving a known treated wastewater discharge into an urban river. Over 16 months, nine campaigns monitored the WWTP discharge and river (upstream and

Table 6

Typical contributors for UV-Vis absorbance spectra and differential spectra interpretation (Guo et al., 2020; Thomas and Burgess, 2022).

Wavelength (nm)	Typical contributors for absorbance spectra		Differential spectra (DS – UP)	Interpretation
	River water	Urban WWTP effluent		
200–230	Nitrate, nitrite, low-molecular-weight DOM	Nitrate, nitrite, organic acids, residual oxidants	Valley at 200–220 nm	Indicates nitrate/nitrite enrichment due to WWTP discharge
220–250	Natural DOM with conjugated double bonds, unsaturated aldehydes and ketones	DOM with conjugated dienes, unsaturated carbonyl compounds, surfactants	Broadening of absorption; sometimes shoulder formation	Suggests addition of organic compounds with conjugated systems from effluent
250–280	Aromatic and humic-like substances of natural origin	Aromatic compounds, phenols, lignin derivatives, dyes	Increased absorbance or shoulder at 260–270 nm	Reflects aromatic and humic substances from treated wastewater
280–380	Humic and fulvic acids, colored DOM	Colored DOM, Fe^{3+} complexes, residual micropollutants	Enhanced peak around 280–300 nm	Indicates input of colored DOM and metal–organic complexes
380–750	Turbidity, Fe^{3+} , Mn, organic colloids	Turbidity, suspended solids, Fe^{3+} /Cr complexes, residual organics	Increased absorbance or slope in visible range	Suggests higher turbidity and metal–organic complex formation downstream

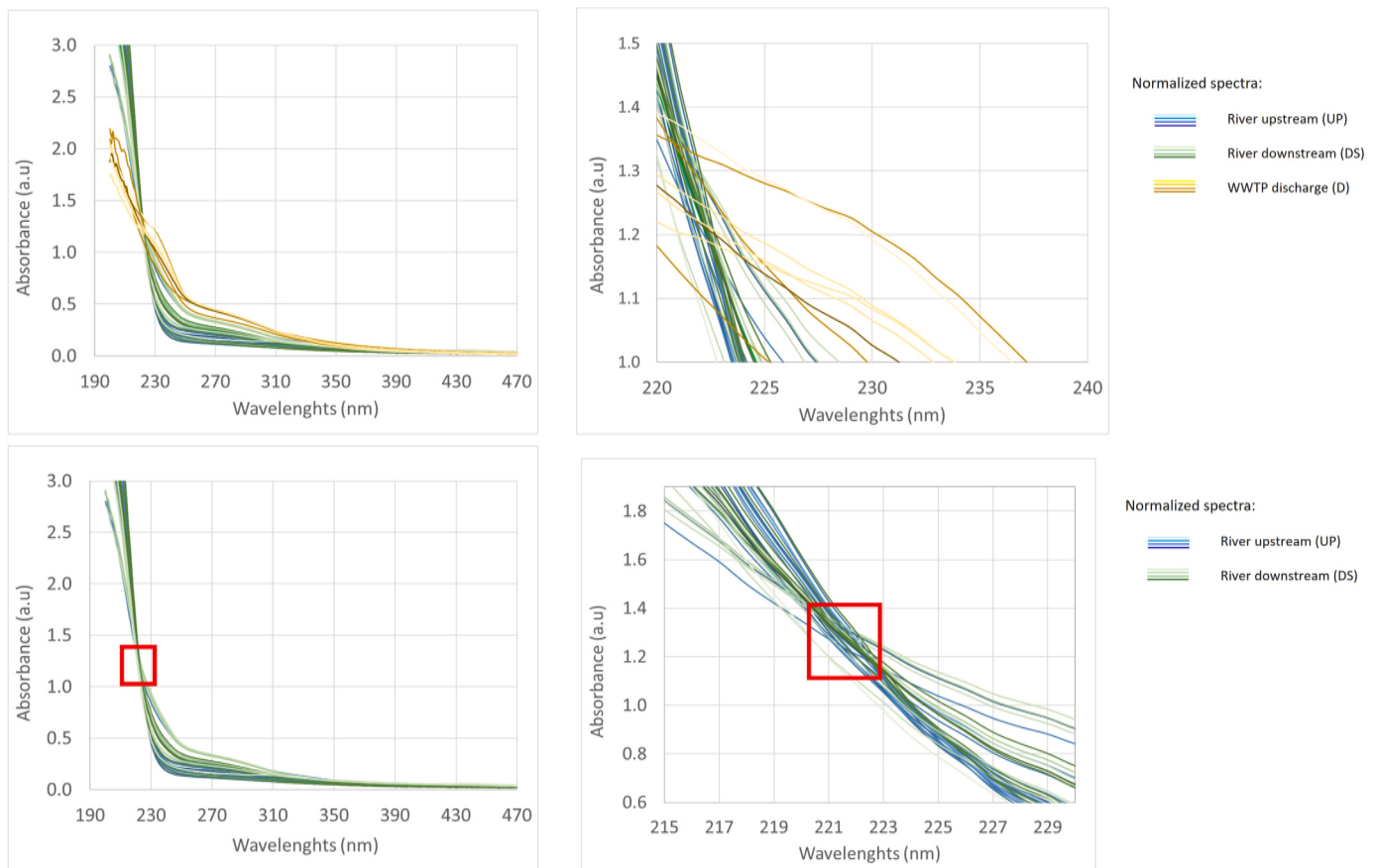


Fig. 8. Normalized spectra from all campaigns (left: 190–470 nm; right: zoom on lower wavelengths). Top: no HIP observed in D, UP and DS spectra. Bottom: HIP appears (red square) when only UP and DS are compared.

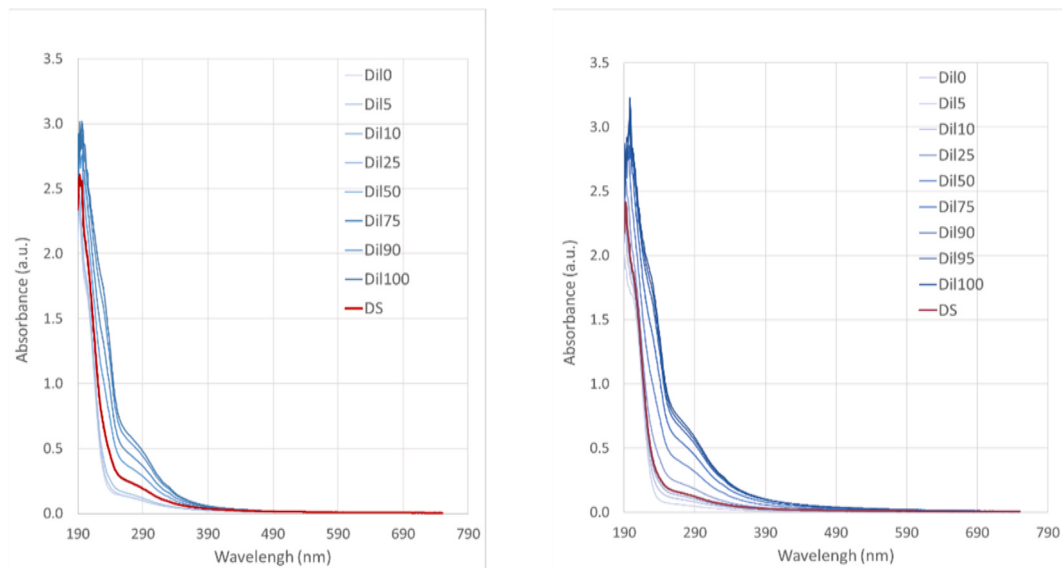


Fig. 9. Positioning of DS spectra among composite Dil_i samples containing $i\%$ D and the remaining UP, for estimating WWTP dilution; campaigns C3 (left) and C5 (right).

downstream), under varying weather and operational conditions. Dry-weather data were more limited, which constrained comparisons with wet conditions, i.e., seasonal variability was not fully captured. Further research on dry-weather data would be required. Out of the nine campaigns, one was conducted without industrial activity, which is relevant

to serve as a baseline to compare with potential episodic discharges. Some campaigns had a limited number of samples per station due to safety conditions and river accessibility. To prevent bias when developing the statistical models, even when more spectra per day were available, only two were used for each river station (UP, DS).

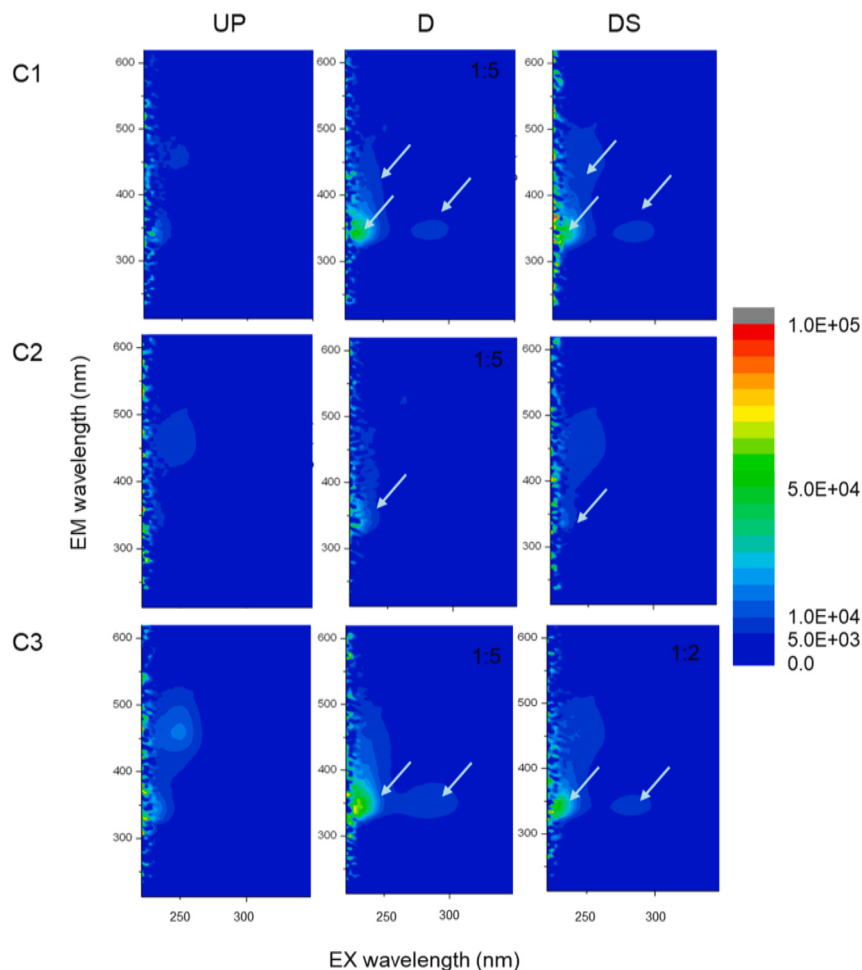


Fig. 10. FEEM contour plots for UP, D and DS samples across campaigns C1-C3; 1:2 and 1:5 refer to dilution ratios with ultrapure water, and color scale represents fluorescence intensity. Arrows indicate common fluorescence regions present in D and DS, but absent or less intense in UP.

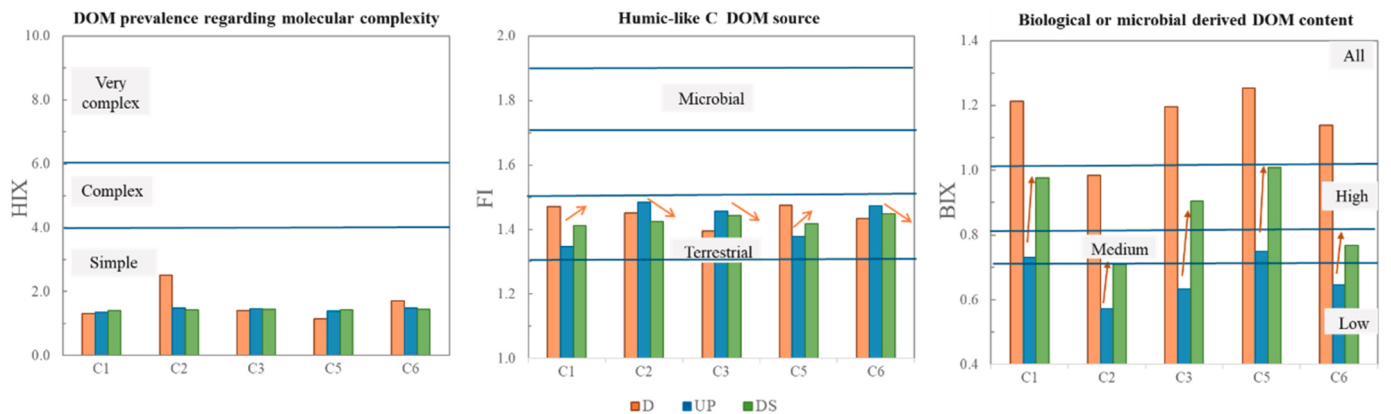


Fig. 11. Fluorescence indices HIX, FI and BIX of D, UP and DS from campaigns C1-C6 (except C4) and relation to DOM sources.

Table 7
Characteristics of PCA models.

Model	n	PC	% Var	RMSE	Outliers
a) Spectra from UP, DS and D	40	3	99.9	0.096	4
b) Spectra from UP and DS	28	3	99.8	0.069	2

n: number of spectra.
PC: required number of principal components.
% Var: percentage of cumulative captured variance.

The method's applicability for *locating and tracing* the discharge was demonstrated (through visual data mining, differential spectrum, iso-sbestic points, PCA or PARAFAC), even when discharged water quality met legal standards, and the river showed prior impairment. As mentioned, the water quality of the local river was considered similar to the effluent of an activated sludge WWTP, based on the Water Quality Label. As the method relies on detecting contrasts between upstream and downstream water quality due to a discharge, it is most effective when the upstream water quality is very different from the discharge, and

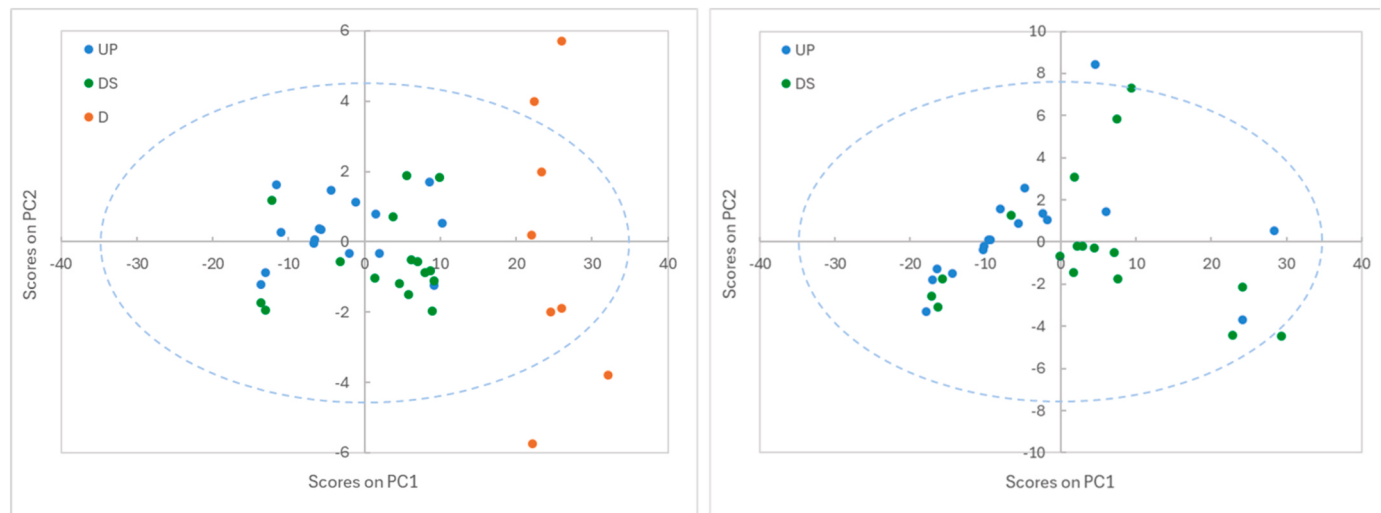


Fig. 12. PCA models scores plots, 95% captured variance in dashed line: (a) UP, DS and D spectra (left); (b) UP and DS spectra (right).

Table 8
Characteristics of PLS models.

Parameters	Range	n	LV	% Var X	% Var Y	RMSE _{rel}	Outliers
TOC (mgC/L)	1.2 – 12.3	28	2	97.2	97.7	14%	None
Conductivity (μS/cm)	88 – 1788	28	3	97.2	95.6	19%	None
COD (mgO ₂ /L)	9 – 47	28	4	99.9	82.1	25%	2
NO ₃ (mg/L)	0.9 – 3.7	28	7	99.9	79.2	33%	3

n: number of spectra.

LV: required number of latent variables.

% Var: percentage of cumulative captured variance.

becomes less effective as this difference decreases, namely due to increasing background pollution in the river. Nevertheless, the method can be generalized to other urban drainage systems with different hydrological or industrial conditions.

The method's applicability for profiling the major water quality components of the discharge was evidenced. Even though many compounds absorb in the same region, simpler procedures such as differential spectra can be valuable to reveal which water constituents are changing between upstream and downstream river samples. The spectral shape is a relevant input. During rain events, dilution generally lowers overall absorbance while preserving the spectral shape; a slight increase may occur in the 300–400 nm region due to terrestrial humic

runoff. In contrast, wastewater discharges increase absorbance around 200–230 nm (nitrite/nitrate and low molecular weight DOM), near 250–280 nm (aromatic and humic-like DOM), and sometimes between 280 and 380 nm (colored DOM). Industrial discharges typically introduce new, often sharper peaks, depending on chemicals (e.g. dyes), altering spectral shape rather than only its intensity.

Discharge dilution was estimated in the case study. PLS-based water quality predictions were partially successful, needing more sampling or complex statistical models. PARAFAC-FEEM enabled quantification of DOM components. Beyond resolving the complexity of water composition, FEEMs enabled the detection of changes in the DOM content of water samples.

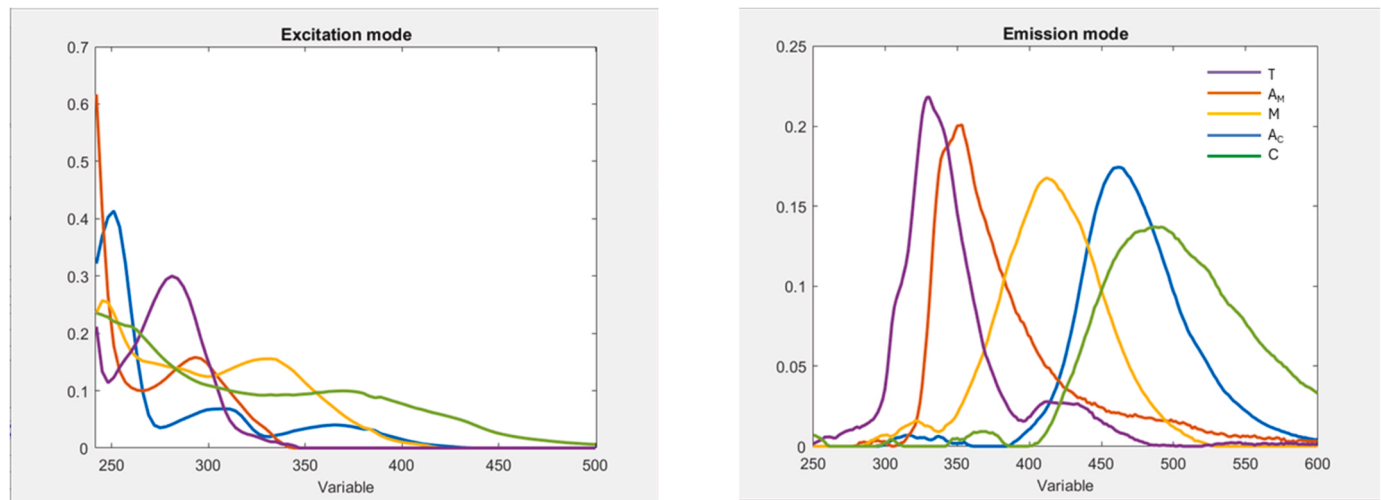


Fig. 13. FEEM PARAFAC components identified at UP, D, DS and association with Coble peaks.

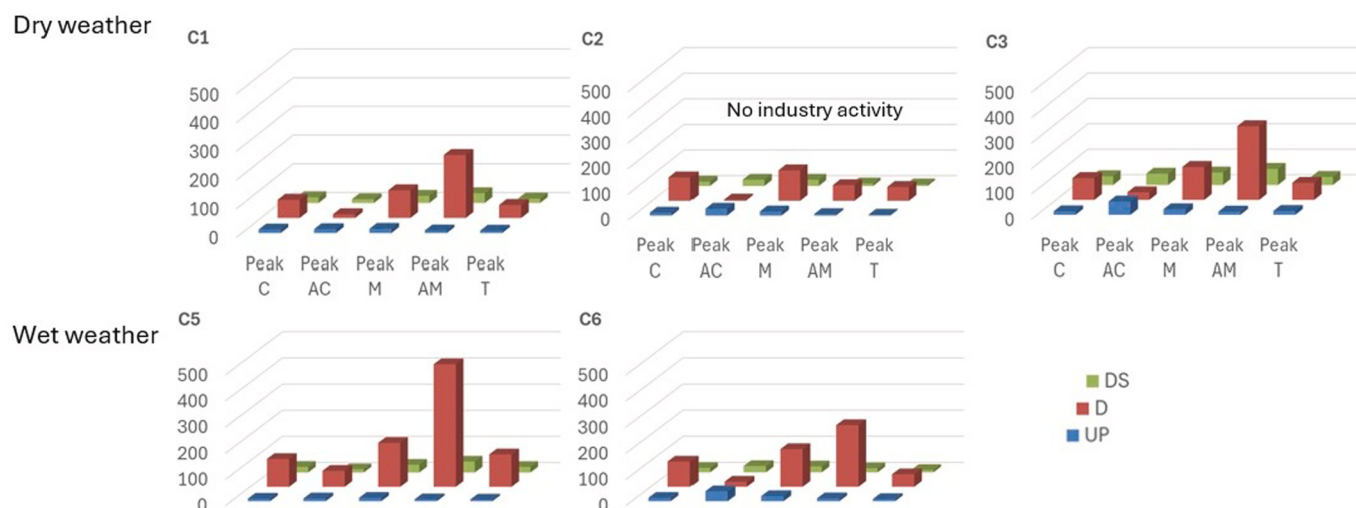


Fig. 14. FEEM PARAFAC component scores (as Coble peaks C, A_C , M, A_M and T) of UP, D and DS samples from campaigns C1-C6).

Table 9

Characteristics of PARAFAC models.

Parameters	Range	n	PARAFAC component (Coble peaks)	% Var Y	RMSE _{rel}
TOC (mgC/L)	2.0 – 6.8	9	A_M	83.3	18%
COD (mgO ₂ /L)	9 – 41	10	T	90.8	11%
Total phosphorus (mg/L)	0.10 – 1.10	10	C	90.9	23%
Total nitrogen (mg/L)	2.3 – 10.2	6	A_M	98.3	8%
Conductivity (μ S/cm)	291 – 1500	10	M	93.1	15%
<i>E. coli</i> (NMP/100 mL)	4.0×10^1 – 9.0×10^3	7	A_C	77.4	14%
Hydrochlorothiazide (ng/L)	11 – 31	6	A_M	97.5	8%

n: number of spectra.

% Var: percentage of cumulative captured variance.

In-situ spectra from three campaigns enabled faster, more detailed river cross-section characterization. Visual data mining, isosbestic point analysis, and PCA of in-situ spectra yielded conclusions consistent with ex-situ measurements. Although online sensors often require higher initial investment, they can reduce sampling and handling errors, lower operational costs, capture short-term fluctuations, provide early pollution warnings, and support WWTP and sewer control. Online water quality monitoring provides continuous, high-frequency data that may capture diurnal and seasonal variability, enabling the distinction between natural fluctuations and anomalous signals, thereby improving the detection of contamination events. While it cannot fully replace the laboratory accuracy, online spectral monitoring offers a highly promising and operationally valuable complement. Simpler online monitoring (of e.g. temperature, turbidity, electrical conductivity or pH) is also particularly useful as a first attempt to quickly detect changes in water quality, making them well-suited for detecting pollution events against a background of natural variability.

Limited knowledge of wastewater discharges creates critical blind spots in urban water management, undermining both compliance and environmental protection efforts. Advanced monitoring strategies are urgently needed to identify, characterize, and manage these discharges effectively – before their cumulative impact grows unmanageable. The proposed methodology can contribute to the swift action mandated by the new UWWTD, protecting public health, safeguarding ecosystems, and ensuring compliance with legally binding EU targets. The timeframe for intervention is limited and becoming closer.

CRedit authorship contribution statement

Rita Salgado Brito: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal

analysis, Data curation, Conceptualization. **Elsa Mesquita:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rui M.C. Viegas:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Joana Soares:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Cláudio Costa:** Writing – review & editing, Resources, Data curation. **Maria João Rosa:** Writing – review & editing, Project administration, Funding acquisition.

Declarations

All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors.

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

Support is acknowledged from the European Union LIFE Fitting project. The publication reflects only the authors' views, and the European Union is not liable for any use made of the information contained herein.

Funding

Funding was received from the European Union LIFE Fitting project: LIFE program, 101,114,188 LIFE22-ENV-PT-LIFE Fitting (<https://life-fitting.lnec.pt/>).

Declaration of competing interest

The authors declare the following financial interests/personal relationships:

Funding was received from the European Union LIFE Fitting project: LIFE program, 101,114,188 LIFE22-ENV-PT-LIFE Fitting (<https://lifefitting.lnec.pt/>).

Acknowledgements

The authors acknowledge technical support by TRATAVE staff, namely Gabriela Faria and Raquel Araújo, and LNEC's laboratory support by Vitor Napier.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jenvman.2026.129759>.

Data availability

Public data used is reported in LIFE Fitting project deliverables

References

- AGI, 2026. *Guidelines for Fresh & Marine Water Quality - Australia & New Zealand*. Australian Government Initiative. <https://www.waterquality.gov.au/anz-guidelines>. (Accessed 27 March 2026).
- Assaad, A., Pontvianne, S., Pons, M., 2017. Assessment of organic pollution of an industrial river by synchronous fluorescence and UV-vis spectroscopy: the Fensch River (NE France). *Environ. Monit. Assess.* 189–229. <https://doi.org/10.1007/s10661-017-5933-3>, 2017.
- Brito, R.S., Ferreira, F., Almeida, J., 2017. Identificação De Águas Residuais Industriais em Coletores Com Recurso a Espectros UV-Vis. *ENEG, Évora, Portugal*, p. 17, 2017 In Portuguese.
- Brito, R.S., Ferreira, F., Pires, R., 2022. Monitoring water quality in urban streams using spectrophotometry. *IWA Specialist Conference LESAM 2022*. Bordéus, França, p. 8.
- Brito, R.S., Lourenço, N., Matos, J.S., Pinheiro, H., 2016. Calibration transfer between spectrometers for wastewater quality monitoring in drainage systems. *Appl. Spectrosc.* 70 (3), 443–454. <https://doi.org/10.1177/0003702815626668>.
- Brito, R.S., Pinheiro, H., Ferreira, F., Matos, J.S., Lourenço, N., 2013. In-situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems. *Urban Water J.* 11 (4), 261–273.
- Booksh, K.S., 2000. In: Meyers, R.A. (Ed.), *Chemometric Methods in Process Analysis, Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Chichester, UK, p. 8145.
- Carstea, E.M., Popa, C.L., Baker, A., Bridgeman, J., 2020. In-situ fluorescence measurements of dissolved organic matter: a review. *Sci. Total Environ.* 699, 134361. <https://doi.org/10.1016/j.scitotenv.2019.134361>.
- Christensen, J., Norgaard, L., Bro, R., Engelsen, S.B., 2006. Multivariate autofluorescence of intact food systems. *Chem. Rev.* 106 (6), 1979–1994. <https://doi.org/10.1021/cr050019q>.
- Carstea, E. M. Bridgeman, J. Baker, A. Reynolds, D. M. 2016 Fluorescence spectroscopy for wastewater monitoring: A review, *Water Research*, 95, 205-219. <https://doi.org/10.1016/j.watres.2016.03.021>.
- Causse, J., Olivier Thomas, O., Jung, A.V., Thomas, M.F., 2017. Direct DOC and nitrate determination in water using dual pathlength and second derivative UV spectrophotometry. *Water Res.* 108, 312–319. <https://doi.org/10.1016/j.watres.2016.11.010>.
- Chen, X., Yin, G., Zhao, N., Gan, T., Yang, R., Xia, M., Feng, C., Chen, Y., Huang, Y., 2021. Simultaneous determination of nitrate, chemical oxygen demand and turbidity in water based on UV-Vis absorption spectrometry combined with interval analysis. *Spectrochim. Acta Mol. Biomol. Spectrosc.* 244, 118827. <https://doi.org/10.1016/j.saa.2020.118827>.
- Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* 51 (4), 325–346.
- Coble, P.G., 2007. Marine optical biogeochemistry: the chemistry of ocean color. *Chem. Rev.* 107 (2), 402–418.
- Coombes, K., Baggerly, K., Morris, J., 2007. Pre-processing mass spectrometry data. *Fundamentals of data mining in genomics and proteomics* 4, 79–107.
- Delpia, I., 2011. *Changements Climatiques Et Dégradation De La Qualité Des Eaux De Surface Destinées À La Consommation Humaine En Zone Agricole*. University of Rennes, France. PhD Thesis.
- Directive (EU) 2024/3019, European Parliament, *Urban Wastewater Treatment Directive* Recast.
- Directive 2008/105/EC, European Parliament, *Environmental Quality Standards Directive*.
- Ma, Y., Li, S., 2020. Spatial and temporal comparisons of dissolved organic matter in river systems of the three gorges reservoir region using fluorescence and UV-Visible spectroscopy. *Environ. Res.* 189, 109925. <https://doi.org/10.1016/j.envres.2020.109925>.
- McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T., 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* 46 (1), 38–48.
- Miller, C., 2000. Chemometrics for on-line spectroscopy applications - theory and practice. *J. Chemometr.* 14, 513–528.
- Murphy, K.R., Stedmon, C.A., Graeber, D., Bro, R., 2013. Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Anal. Methods* 5 (2), 6557–6566. <https://doi.org/10.1039/c3ay41160e>.
- Oliveira, M., Levkowitz, H., 2003. From visual data exploration to visual data mining: a survey. *IEEE Trans. Visual. Comput. Graph.* 9, 378–394. <https://doi.org/10.1109/TVCG.2003.1207445>.
- Platikanov, S., Rodríguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., Tauler, R., 2014. Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. *J. Environ. Manag.* 140, 33–44. <https://doi.org/10.1016/j.jenvman.2014.03.006>.
- Pouet, M., Baurès, E., Vaillant, S., Thomas, O., 2004. Hidden isosbestic point(s) in UV spectra. *Appl. Spectrosc.* 58, 486–490.
- Pu, H., Liu, J.-H., Qu, D., Sun, D.-W., 2017. Applications of imaging spectrometry in inland water quality monitoring—a review of recent development. *Water Air Soil Pollut.* 228, 131. <https://doi.org/10.1007/s11270-017-3294-8>.
- Rice, E.W., Baird, R.B., Eaton, A.D., Clesceri, L.S., 2012. *Standard Methods for the Examination of Water and Wastewater*, twenty-second ed. American Public Health Association, American Water works Association, Water Environment Federation.
- Ritson, J.P., Kennedy-Blundell, O., Croft, J., Templeton, M., Hawkins, C., Clark, J., Evans, M., Brazier, R., Smith, D., Graham, N., 2022. High frequency UV-Vis sensors estimate error in riverine dissolved organic carbon load estimates from grab sampling. *Environ. Monit. Assess.* 194, 831. <https://doi.org/10.1007/s10661-022-10515-9>.
- Rong, Y., Wang, X., Zhu, M., Chen, W., Maletskyi, Z., Ratnaweera, H., Bi, X., 2024. Real-time ozonation monitoring and control in wastewater tertiary treatment: an ultraviolet-visible monitoring approach. *J. Water Proc. Eng.* 64, 105664. <https://doi.org/10.1016/j.jwpe.2024.105664>.
- Serène, L., Mazzilli, N., Batiot-Guilhe, C., Emblanch, C., Gillon, M., Babic, M., Dupont, J., Simler, R., Blanc, M., 2025. Questioning calculation and interpretation of fluorescence indices in natural waters studies. *J. Hydrol.* 650, 132524.
- Shi, Z., Chow, C.W.K., Fabris, R., Liu, J., Jin, B., 2022. Applications of online UV-Vis spectrophotometer for drinking water quality monitoring and process control: a review. *Sensors* 22 (8), 2987. <https://doi.org/10.3390/s22082987>.
- Soares, J., 2025. *Application of Fluorescence Spectroscopy for Expeditious Characterization of Treated Wastewater for Reuse*. PhD thesis. Instituto Superior Técnico, Lisbon, Portugal.
- Stedmon, C., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* 82 (3–4), 239–254.
- Stedmon, C.A., Bro, R., 2008. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnol. Oceanogr. Methods* 6 (11), 572–579. <https://doi.org/10.4319/Jom.2008.6.572>.

- Thomas, O., Burgess, C., 2022. UV-Visible Spectrophotometry of Waters and Soils, third ed. Elsevier. <https://doi.org/10.1016/B978-0-323-90994-5.00017-4>. ISBN 9780323909945.
- Thomas, O. 1995. Metrologie des eaux residuaires, Lavoisier Tech. et Doc., Paris, Liege. 33_34.
- Torres, A., Bertrand-Krajewski, J.L., 2008. Partial least squares local calibration of a UV-visible spectrometer used for in situ measurements of COD and TSS concentrations in urban drainage systems. *Water Sci. Technol.* 57 (4), 581–588.
- Van der Broeke, J., Langergraber, G., Wiengartner, A., 2006. On-line and in-situ UV/vis spectroscopy for multi-parameter measurements: a brief review. *Spectrosc. Eur.* 18 (4).
- Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Sci. Technol.* 47 (2), 1–34.
- Viegas, R.M.C., Mestre, A.S., Mesquita, E., Machuqueiro, M., Andrade, M.A., Carvalho, A. P., Rosa, M.J., 2022. Key factors for activated carbon adsorption of pharmaceutical compounds from wastewaters: a multivariate modelling approach. *Water* 14 (2), 166. <https://doi.org/10.3390/w14020166>.
- Yang, L., Hur, J., Zhuang, W., 2015. Occurrence and behaviors of fluorescence EEM PARAFAC components in drinking water and wastewater treatment systems and their applications: a review. *Environ. Sci. Pollut. Res.* 1–11.
- Zhang, H., Zhang, L., Wang, S., Zhang, L., 2022. Online water quality monitoring based on UV-Vis spectrometry and artificial neural networks in a river confluence near sherfield-on-loddon. *Environ. Monit. Assess.* 194, 630. <https://doi.org/10.1007/s10661-022-10118-4>.
- Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to dry. *Chemosphere* 38, 45–50. [https://doi.org/10.1016/S0045-6535\(98\)00166-0](https://doi.org/10.1016/S0045-6535(98)00166-0).