Petroleum Science 23 (2026) 2216–2230

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Heterogeneity in the molecular composition of effluent organic matter from petrochemical wastewater treatment plants in China

Yue Kou [a,1], Huang-Fan Ye [b,1], Rui Zhang [c], Wen-Yi Lu [c], Ze-Lin Niu [c], Yi-Fei Zhu [c], Zhuo-Yu Li [c,*], Chen He [c], Hai-Feng Wang [a,**], Qing-Hong Wang [c], Quan Shi [c], Chun-Mao Chen [c]

[a] National Institute of Metrology of China, Beijing, 100029, China
[b] College of Chemistry and Chemical Engineering, Chongqing University of Science and Technology, Chongqing, 401331, China
[c] State Key Laboratory of Heavy Oil Processing, State Key Laboratory of Petroleum Pollution Control, China University of Petroleum (Beijing), Beijing, 102249, China

ARTICLE INFO

Article history:
Received 1 August 2025
Received in revised form
18 December 2025
Accepted 10 March 2026
Available online 18 March 2026

Edited by Min Li

Keywords:
Effluent organic matter
Petrochemical wastewater
Molecular characteristic
Heterogeneity
Orbitrap mass spectrometry

ABSTRACT

The proliferation of petrochemical plants close to natural water bodies has resulted in emission of numerous organic pollutants into aquatic environments. However, the molecular composition of petrochemical effluent organic matter (PEfOM) is poorly understood, which results in uncertainty regarding its environmental effects and potential risks. In this study, optical spectroscopy and mass spectrometry coupled with both negative-ion ((−)ESI mode) and positive-ion electrospray ionization mode ((+)ESI mode) were used to conduct large-scale molecular-level characterization of PEfOM from 21 petrochemical plants in China. In contrast to natural organic matter (NOM), PEfOM was characterized by low aromaticity and high microbial-derived fluorescence. These properties corresponded to a composition rich in aliphatic and highly unsaturated phenolic compounds, as well as low average molecular weight and elevated heteroatom content. Among the 21 PEfOM samples, molecular compositions exhibited significant heterogeneity with Jensen–Shannon divergence values between 0 and 0.2. Based on the O/S/N contents, aromaticity, and unsaturation, the compounds in PEfOM were grouped into three clusters under (+)ESI mode (+CL1, +CL2, +CL3) and two clusters under (−)ESI mode (−CL1 and −CL2). Nitrogen-rich CHON compounds with high double bond equivalents (4–14) in +CL2 and oxygen-poor compounds with high molecular weights (600–800 Da) in −CL1 exhibited the greatest reactivity in aquatic environments under aerobic conditions. The compounds in these (+)ESI and (−)ESI clusters had Gibbs's free energy of carbon oxidation values of 80–110 and 60–100, respectively. $C_{17}H_{23}N_1O_2$ (quinolinium derivatives), $C_{14}H_{20}O_5S_1$ (benzene sulfonic acid), and $C_9H_{17}N_1O_5S_1$ (amino-sulfonates) were identified as characteristic pollutants with high indicator species values. Ecological structure activity relationships analysis revealed that $C_{17}H_{23}N_1O_2$ exhibited high acute biotoxicity, which was associated with potential environmental risks. This study offers valuable insight and guidance for risk management and control of petrochemical effluents.

1. Introduction

With development of the petrochemical industry, the number of petrochemical plants located in riverine and coastal areas has gradually increased and the production and discharge of petrochemical wastewater from these plants has attracted attention (Mansour et al., 2024). Because of the complex reactions in the petrochemical industry (Fatkullina et al., 1979), wastewater generated from petrochemical facilities usually contains hundreds

* Corresponding author.
** Corresponding author.
E-mail addresses: lizhuoyu@cup.edu.cn (Z.-Y. Li), wanghf@nim.ac.cn (H.-F. Wang).

Peer review under the responsibility of China University of Petroleum (Beijing).

[1] Authors contributed equally.

https://doi.org/10.1016/j.petsci.2026.03.021

of organic pollutants produced from raw materials, auxiliary materials, and reaction by-products (Li et al., 2024a). Compared with natural organic matter (NOM) in aquatic environments (e.g. humic acids and fulvic acids), the DOM in petrochemical wastewater from different treatment processes has a lower molecular weight, higher degree of unsaturation, and contains more complex aromatic structures (He et al., 2023). To meet environmental regulations, petrochemical wastewater usually undergoes a series of intricate physiochemical and biochemical treatment processes before being discharged. Although the pollution load of the petrochemical wastewater can be greatly reduced by diverse microbial metabolic pathways and chemical reactions within treatment units, many organic molecules are also generated that further increase the complexity of DOM components in petrochemical wastewater. To date, research has focused on reduction of pollution loads during petrochemical wastewater treatment processes. The bulk indicators of chemical oxygen demand, total organic carbon content, and biochemical oxygen demand have been used to characterize petrochemical effluent quality. However, the specific molecular composition of petrochemical effluent organic matter (PEfOM) and how it differs from NOM are not clear. The unique composition of PEfOM likely governs its subsequent transformation pathways and environmental effects. After discharge into the aquatic environment, PEfOM inevitably undergoes a series of photochemical or biodegradation transformation processes, which is likely to produce compounds with adsorption capacities, reactivities, and toxicities that differ from those of NOM. These compounds may have the potential to disrupt the chemical equilibrium and ecological processes in natural water bodies. Previous studies have shown that photochemical reactions shorten carbon chains in effluent organic matter in municipal wastewater, and the products potentially act as substrates for microbial growth and lead to oxygen depletion in water bodies (Zheng et al., 2025). Sunlight-induced transformation of effluent organic matter may also lead to the formation of haloketones, which are toxic and persistent in the environment (Du et al., 2023). The complex composition of PEfOM creates uncertainty about its environmental transformation and potential risks. A systematic analysis of the molecular composition and characteristics of PEfOM is vital for investigating its potential reactivity and environmental effects.

To complement the commonly used bulk indicators, recently developed optical spectroscopy and mass spectrometry techniques are feasible for comprehensive characterization of DOM components. Ultraviolet-visible (UV-vis) spectroscopy has been combined with three-dimensional excitation–emission matrix (3D-EEM) spectroscopy to elucidate the aromaticity of DOM and features of fluorescent components (Phungsai et al., 2016; Wang et al., 2015). While optical spectroscopy cannot gain comprehensive characterization of DOM components at the molecular level. High-resolution mass spectrometry, especially Orbitrap mass spectrometry (MS), can be used to compensate for the limitations of conventional gas chromatography-MS and is a powerful tool for investigating the molecular structures of polar DOM with excellent resolution and accuracy (Li et al., 2024c). Optical spectroscopy and Orbitrap MS have been successfully combined to characterize the molecular compositions of DOM in various complex water matrices, including landfill leachate (Bao et al., 2024; Nguyen et al., 2024), drinking water (Phungsai et al., 2018), municipal wastewater (Zhang et al., 2021), NOM (Hawkes et al., 2016), and refinery wastewater (Li et al., 2024b; Vanini et al., 2020). The choice of electrospray ionization (ESI) modes for Orbitrap MS is crucial for acquiring comprehensive molecular composition information from DOM samples. Zhao et al. (2023) discovered that DOM molecules containing acidic functional groups were identifiable by negative ESI ((−)ESI) mode because of their ease of ionization and deprotonation. By contrast, unsaturated hydrocarbons and nitrogen-containing compounds are readily protonated and ionized in positive ESI ((+)ESI) mode (Laszakovits et al., 2020). Given the high molecular complexity of petrochemical effluents, both (+) and (−)ESI modes have been used to simultaneously detect acidic (e.g., sulfonates) and basic (e.g., nitrogenous heterocycles) compounds (Li et al., 2020). This combination of modes allows for accurate molecular profiling of PEfOM.

In addition to optimizing analytical techniques to obtain more detailed information on the molecular composition of PEfOM, revealing the underlying characteristics and heterogeneity is equally important. Petrochemical wastewater samples generated by different production units have distinct DOM compositions because of differences in the reaction processes. According to our previous research, DOM from crude oil electric desalting wastewater is dominated by aromatic amines, alkylbenzene sulfonates, and naphthenic acids, whereas phenol–acetone plant wastewater mainly contains oxygenated compounds such as ketones and aromatic acids (Kou et al., 2023b). Therefore, changes in production processes and crude oil source among various petrochemical plants may lead to different compositions of petrochemical wastewater and variation in PEfOM components. The extensive physical, chemical, and biological treatments during both petrochemical wastewater treatment processes are hypothesized to be the dominant factors shaping the final molecular signature of PEfOM, as these processes can introduce new, treatment-specific compounds.

Elucidation of common characteristics among PEfOM samples and the level of heterogeneity requires analysis of many effluent samples and the use of statistical methods to identify patterns and differences. Cluster analysis has been used to investigate the heterogeneity and common characteristics of highly connected natural water body samples and ground water samples over large areas (Cai and Jiao, 2023; Harjung et al., 2023). However, few studies have conducted large-scale sampling to comprehensively characterize PEfOM.

The main objectives of this study were to (1) conduct a large-scale analysis of the molecular composition of PEfOM from 21 petrochemical plants across China, (2) identify the common molecular characteristics of PEfOM and elucidate its compositional differences from NOM, and (3) identify structures and acute biotoxicities of representative compounds in PEfOM. The results provide detailed insight into the molecular-level composition of PEfOM, which is crucial for understanding its environmental effects.

2. Materials and methods

2.1. Sample collection and preparation

PEfOM samples were collected from 21 petrochemical wastewater treatment plants across China between February and March 2023. The plants were located along rivers and oceans in different regions. Ten plants were located along the coastline of China (samples #01–#10), six plants were located along the Yellow River (samples #11–#16), and five plants were located along the Yangtze River (samples #17–#21) (Fig. S1). The types of plants and the properties of their oil products (density, sulfur content, and acid value) varied among the plants (Table S1). To analyze the composition differences between PEfOM and NOM samples, we used NOM samples extracted from the South China Sea, a lake in North China, and a river in North China as representative background benchmarks for major natural water types, which the industrial signature of PEfOM could be contrasted against. The

standard Suwannee River NOM (SRNOM) was purchased from the International Humic Substances Society.

All of the PEfOM and NOM samples were collected in duplicate using brown polypropylene bottles, which were prewashed with Milli-Q water to minimize contamination. The samples were filtered through 0.45-μm glass fiber membrane filters (Tianjin Jinteng Experimental Equipment Co., Ltd., Tianjin, China) to remove particulate matter, stored at 4 °C, and analyzed within 1 week. During the experimental analysis, the dissolved organic carbon (DOC) concentration of the SRNOM standard solution was adjusted to 10 mg/L to match the typical DOC concentrations of the PEfOM samples.

2.2. Analysis of physicochemical and optical indicators

The bulk indicators, DOC, total dissolved nitrogen, nitrate nitrogen, nitrite nitrogen, electrical conductivity, and dissolved organic nitrogen, were measured in triplicate as described in our previous report (Kou et al., 2023a). The optical properties of PEfOM were characterized using a UV-vis spectrometer (UVmini-1280, Shimadzu, Tokyo, Japan) and 3D-EEM as detailed in our previous study (Kou et al., 2023a). To mitigate potential interference from inner filter effects and fluorescence quenching, the UV absorbance at 254 nm for each sample was diluted below 0.3 cm^{-1} using Milli-Q water before 3D-EEM analysis. Parallel factor analysis was used to decompose the intricate fluorescent DOM into distinct fluorescent components in 3D-EEM data (Wünsch et al., 2017). The maximum fluorescence intensity for each component was used as a proportional indicator of its relative concentration (Goldman et al., 2012).

2.3. Analysis of PEfOM molecular compositions

To identify nitrogen-, oxygen-, and sulfur-containing compounds in PEfOM, an Orbitrap fusion mass spectrometer (Thermo Fisher Scientific, Waltham, MA USA) with a resolving power of 500,000 at *m/z* 400 was used to acquire the primary full scan mass spectra of PEfOM under (+)ESI and (−)ESI modes at *m/z* 100–800. Solid-phase extraction of PEfOM was accomplished using methanol, followed by its direct infusion into the ESI source at a flow rate of 10 μL/min. Molecular formulas were assigned using the elemental composition limits of $^{12}C_{1–50}$, $^{1}H_{1–80}$, $^{16}O_{0–30}$, $^{14}N_{0–4}$, $^{32}S_{0–2}$, and $^{13}C_{0–2}$, with a mass accuracy tolerance of ±1 ppm. The molecular weight (MW), hydrogen/carbon ratio (H/C), oxygen/carbon ratio (O/C), nominal oxidation state of carbon (NOSC), nitrogen/carbon ratio (N/C), double bond equivalents (DBE), and Gibbs's free energy of carbon oxidation (ΔG_0) were calculated as weighted averages of the peak intensities from the full scan mass spectra (Liao et al., 2024; Shao et al., 2021).

Orbitrap tandem mass spectrometry (MS/MS) analyses were conducted using Thermo Xcalibur Qual Browser. Collision induced dissociation was used as the fragmentation technique in MS/MS, and selected peaks were targeted for secondary mass spectra acquisition. The complete angle scatter instrument window was precisely adjusted to encompass the *m/z* range of interest, with a tolerance of ±0.5 ppm. To identify the chemical structures of molecules at specific *m/z* values, MetFrag was used to retrieve potential candidate structure from the PubChem database (Ruttkies et al., 2016, 2019b, 2019a). The molecular structures of compounds were inferred using the matching coefficients (Finalscore) of fragment ion peaks with standard spectra and the number of matching peaks. A Finalscore of 1 or the highest number of matching peaks showed the compound's structure closely resembled the true structure. Each compound was potentially associated with more than one candidate structure because the structure with the highest Finalscore did not always coincide with the one exhibiting the most fragment peak matches. To communicate the identification confidence, we adopted the (Schymanski et al., 2014) framework. A Level 1 identification (confirmed structure) requires a matching authentic standard, which was not available for these novel industrial compounds. Identifications supported by a library spectrum match and diagnostic fragment ions are classified as Level 2 (probostic structure). When multiple candidate structures exist and the assignment is based on in-silico fragmentation without a reference spectrum, it is classified as Level 3 (tentative candidate). It is important to note that Levels 2 and 3 identifications, while highly informative for characterizing complex mixtures, retain a degree of uncertainty as they are based on computational matching and cannot definitively distinguish between isomeric structures without further orthogonal evidence.

2.4. Statistical analysis of PEfOM composition differences

The molecular distribution of PEfOM was analyzed using a combination of rank abundance curves and species accumulation curves (95% confidence intervals), which were generated from the normalized relative intensities (RIs) of all compounds using the biodiversity and vegan packages in R, respectively. Jensen–Shannon divergence (JSd) was used to quantitatively assess PEfOM molecular dissimilarities among samples by comparing the frequency distributions of selected molecular attributes (modified aromaticity index (AI_{mod}), DBE, O/C, and H/C) across samples (Lin, 1991). These parameters were selected because they are fundamental molecular descriptors that collectively capture key aspects of DOM chemical diversity and reactivity. Specifically, AI_{mod} and DBE reflect the aromaticity and unsaturation of molecules, which are closely associated with their environmental persistence and photochemical reactivity (Koch and Dittmar 2006). The O/C and H/C inform on the oxidation state and saturation of organic compounds, which influence their polarity, bioavailability, and overall environmental behavior (He et al., 2022). The combination of these parameters allows the JSD analysis to effectively quantify differences in molecular composition based on chemically meaningful properties. Molecular formulas were first classified into four categories: aromatic compounds ($AI_{mod} > 0.5$), highly unsaturated compounds ($AI_{mod} \leq 0.5$, H/C < 1.5), unsaturated compounds (H/C ≥ 1.5, DBE ≠ 0), and saturated compounds (DBE = 0). Subsequently, these four categories were further stratified using O/C. Compounds with O/C < 0.5 were categorized as low-oxygen compounds and those with O/C > 0.5 as high-oxygen compounds. The JSd calculation process used a dedicated function code available on Zenodo (Kida et al., 2023). The JSd values were between 0 and 1, with higher values indicating greater molecular heterogeneity between the PEfOM samples.

The differentiated molecular compositions of PEfOM were identified through *k*-means hierarchical clustering analysis using the calculated JSd values. The fviz_nbclust function in R software (v. 4.2.0) was used to visualize the clustering analysis results, with the optimum number of clusters determined by maximizing the silhouette coefficient. The representative molecular formulas for each cluster were chosen using their indicator species value (IndVal) results (Eqs. (1)–(3)). The IndVal results were ranked from 0 to 1, with higher values indicating a stronger association of a molecule with a cluster. The indval function in R (v. 4.2.0) was used to calculate the IndVal and *P*-value for each molecule in each sample.

$$\mathrm{IndVal}_{ij} = A_{ij} \times B_{ij}, \quad (1)$$

$$A_{ij} = \mathrm{RI}_{ij} \Big/ \mathrm{RI}_i, \tag{2}$$

$$B_{ij} = N_{ij} \Big/ N_j, \tag{3}$$

where A_{ij} is relative abundance of molecule i in cluster j, B_{ij} is relative frequency of molecule i in cluster j, RI_{ij} is relative intensity of molecule i in cluster j, RI_i is relative intensity of molecule i in all clusters, N_{ij} is the number of samples containing molecule i in cluster j, N_j is total number of samples in cluster j, and IndVal_{ij} is IndVal of molecule i in cluster j.

3. Results and discussion

3.1. Physicochemical and spectroscopic properties of PEfOM

3.1.1. Physicochemical properties of PEfOM

The physicochemical and spectroscopic properties of PEfOM samples were investigated using bulk indicators and optical indicators (Fig. 1). The pH (7.52 ± 0.49), DOC concentration (9.95 ± 2.91 mg/L), and total dissolved nitrogen concentration (6.55 ± 1.99 mg/L) exhibited narrow distribution ranges among the 21 samples. The measured DOC and TDN concentration were well below the Chinese Emission standard (Ministry of Ecology and Environment of the People's Republic of China, 2015)'s stipulated limit for Total Organic Carbon (TOC) at 20 mg/L and total nitrogen (TN) at 40 mg/L, respectively. It indicated that PEfOM commonly had low contents of organic and nitrogen-containing pollutants. The dissolved organic nitrogen concentration (3.65 ± 2.34 mg/L) and total dissolved solid (TDS) concentration (4213.14 ± 1588.83 mg/L) showed relatively broad distribution ranges among the PEfOM samples. These results reflected differences in the organic nitrogen content and salinity between samples (Fig. 1(a)).

3.1.2. Optical properties of PEfOM

The specific UV absorption at 254 nm (SUVA_{254}) is a widely used indicator for the aromaticity of DOM, as aromatic compounds (e.g., those with conjugated double bonds) strongly absorb UV light at this wavelength. SUVA_{254} for PEfOM varied widely from 0.5 to 2.8 L/(mg·cm) (Fig. 1(b)), indicating a wide distribution in aromaticity among PEfOM samples. The Fluorescence Index (FI) is used to infer DOM sources: an FI > 1.9 typically indicates a predominantly microbial origin, whereas an FI < 1.7 suggests a terrestrial source. Although the fluorescence index (FI) values also covered a wide range (1.86–2.67), most PEfOM samples had FI values of >1.9. Therefore, the main fluorescent components in PEfOM were endogenous microbial metabolites originating from biological wastewater treatment processes (He et al., 2019). By contrast, NOM samples exhibited much higher SUVA_{254} values of approximately 3 L/(mg·cm), which showed they contained more aromatic compounds originating from industrial processes than the PEfOM samples. The FI values of NOM covered a narrow range (1.75–1.8), which indicated that the dominant compounds in NOM were from exogenous organic matter such as plants and soil. Overall, the PEfOM samples had relatively similar physicochemical properties and low contents of organic and nitrogen-containing pollutants. The optical properties differed between PEfOM and NOM, which suggested the compounds they contained had diverse origins and generation pathways. Interestingly, the optical properties among different PEfOM samples were not as uniform as those among NOM samples, which was related to variations in the source of oil and petrochemical and wastewater treatment processes.

To identify the key fluorescent components in PEfOM, the 3D-EEM spectra of both PEfOM and NOM were analyzed using parallel factor analysis (Murphy et al., 2014). The fluorescence of SRNOM was centered at Ex/Em maxima of 225–275 nm/300–350 and 400–475 nm (Fig. S3). Peaks at these positions are primarily attributed to humic-like compounds (He et al., 2019). For the NOM samples from seawater, lake water, and river water, fluorescence peaks was observed at Ex of 225–250 nm and Em of 275–300 nm and 300–350 nm (Fig. S3), which were attributed to protein-like compounds (He et al., 2019). Given the large sample size and anticipated complexity of the petrochemical effluents, PARAFAC analysis was applied to the entire PEfOM dataset to resolve the underlying common fluorescent components that persist across different plants. In contrast to the diverse but characteristic signatures of the NOM samples, the PARAFAC model identified two robust fluorescent components (C1 and C2) that were ubiquitous in the PEfOM samples (Fig. 1(c)). This comparison between the distinct, characteristic spectra of major NOM types and the dominant, shared components of PEfOM underscores a fundamental difference in the origin and composition of their fluorescent DOM. For C1, fluorescence was centered at Ex of 225–250 and 325–350 nm and Em of 375–425 nm, with the similarity index of over 95% to fluorescent components in four studies in the OpenFluor database. For C2, fluorescence was centered at Ex of 225–250 nm and Em of 325–350 and 375–425 nm, with the similarity index of over 95% to fluorescent components in 21 studies in the OpenFluor database. The higher number of matches for C2 compared with that for C1 indicated that C2 was more prevalent in aquatic environments. To explore the composition of C1 and C2 in detail, Spearman's correlation analysis between the identified fluorescent components and optical indicators of PEfOM was conducted (Fig. S2). Both C1 and C2 exhibited significant positive correlations with SUVA_{254} values ($P < 0.01$), which indicated that aromatic structures were present in both fluorescent components. Moreover, C1 was positively correlated with the FI ($P < 0.01$), which is associated with bio-derivatives for PEfOM. Given that C1 is uncommon in aquatic environments, we inferred that petrochemical effluent would introduce uncommon aromatic bio-derivatives to the receiving water body.

3.2. Molecular features of PEfOM and NOM

To evaluate the representativeness of the PEfOM samples for molecular characterization, the cumulative curve of the Orbitrap MS data was examined (Fig. 2(a) and (b)). More than 15% of all compounds were present in at least 90% of all samples in (+)ESI and (−)ESI modes. These results demonstrated that a core set of DOM compounds were present across a large spatial scale; therefore, the analyzed PEfOM samples were highly representative. Notable saturation in the accumulation of unique molecules was observed in the cumulative curves (Fig. 2(c) and (d)), with 16 out of 21 samples achieving 95% of compound richness in (−)ESI mode and 18 out of 21 samples in (+)ESI mode, signifying the adequate numbers of the PEfOM samples. However, the cumulative curve for PEfOM approached saturation later than for previously reported NOM samples (Kellerman et al., 2014). In the previous study, the cumulative curve approached saturation of molecular accumulation within the first 30% of samples (Kellerman et al., 2014). For NOM, a few samples can adequately reflect its overall characteristics, indicating relative homogeneity across its composition. For PEfOM, however, a much larger sample set is required to capture its generalizable features. This directly demonstrates that the total molecular diversity of PEfOM across China is vastly greater and more spatially variable than that of NOM. For NOM, a few samples can adequately reflect its overall

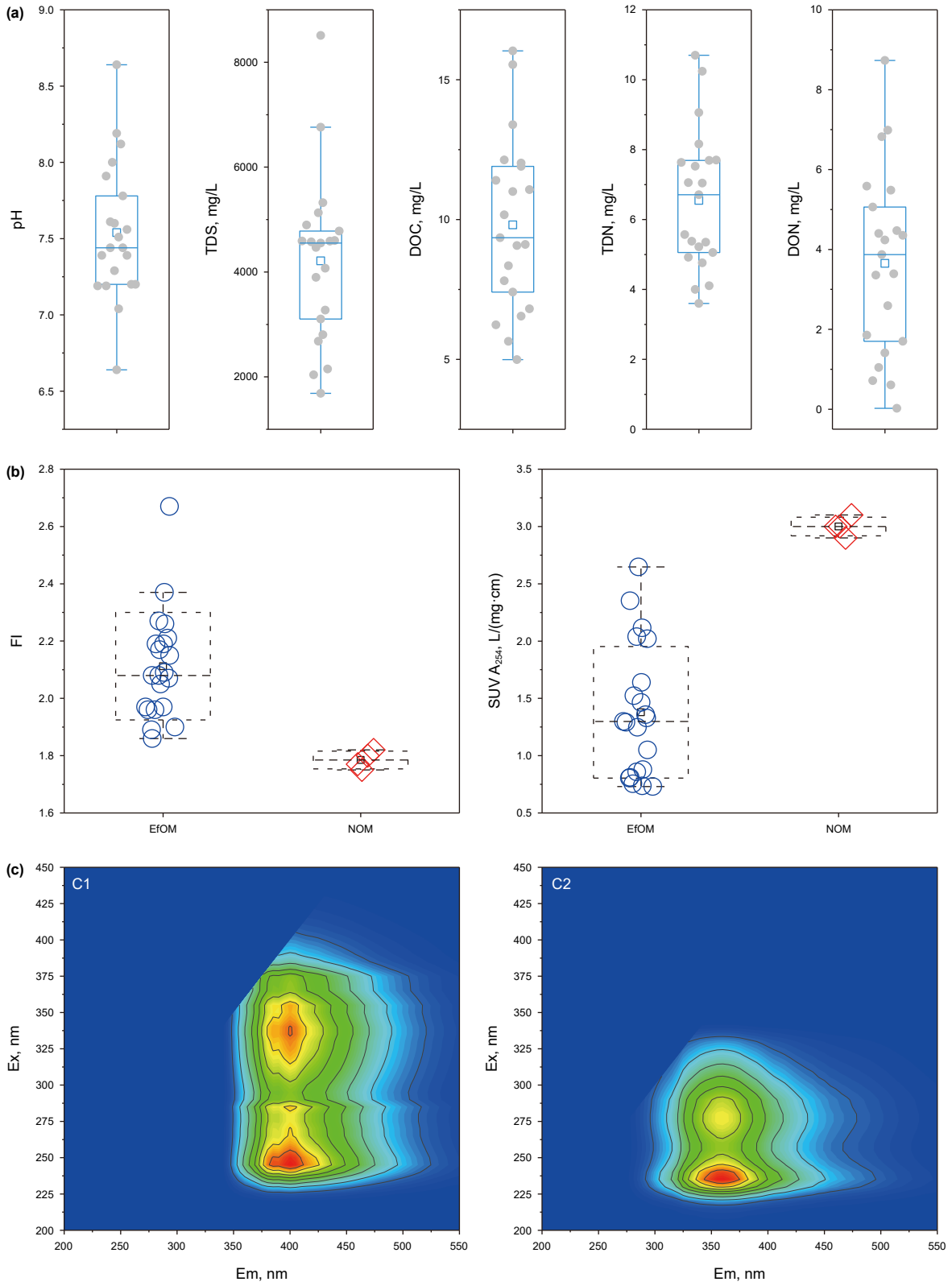


Fig. 1. Optical indicators (**a**) and fluorescence components ((**b**) and (**c**)) in petrochemical effluent organic matter and natural organic matter. Ex: excitation wavelength, Em: emission wavelengths.

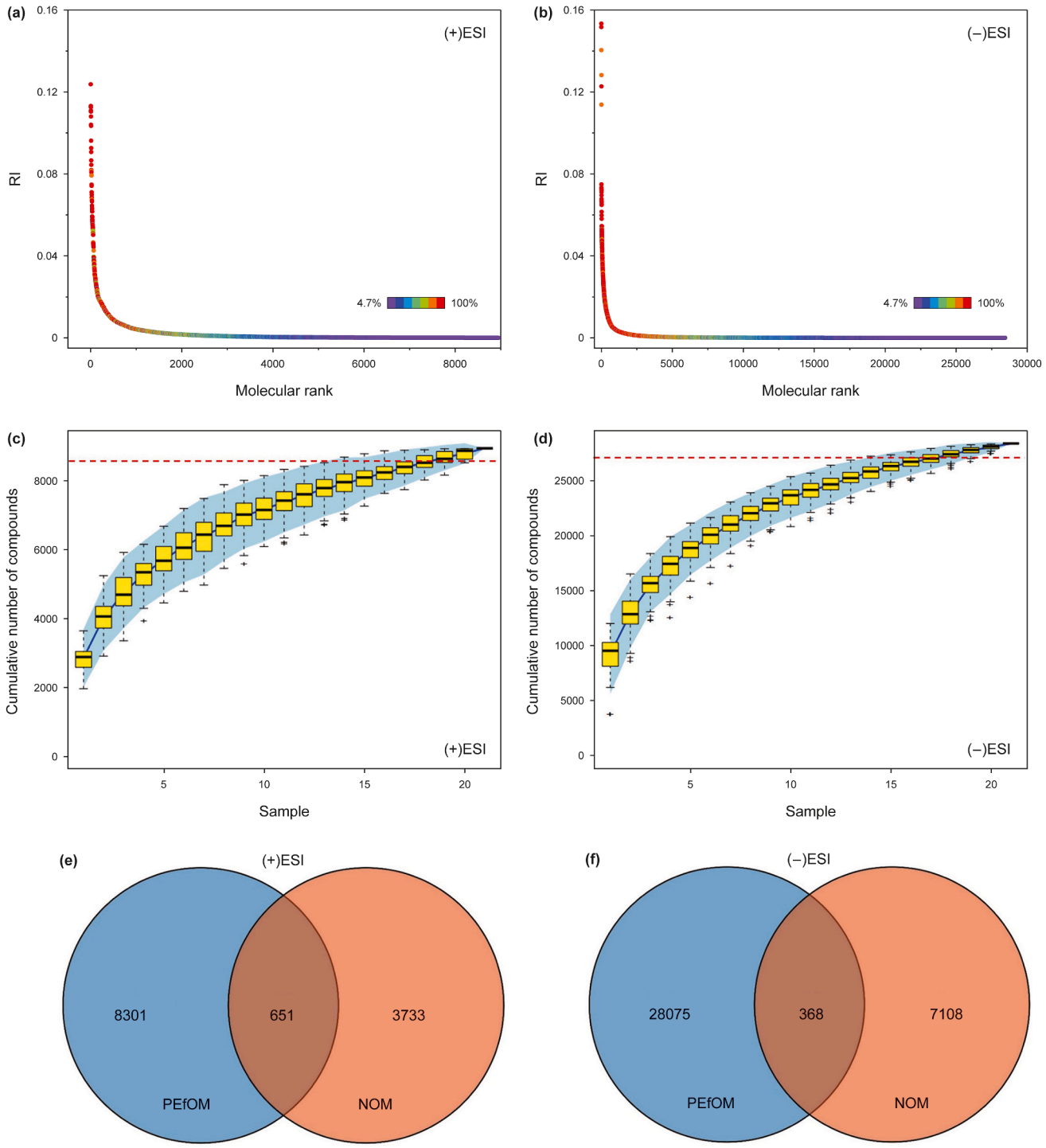


Fig. 2. Rank abundance curves ((**a**) and (**b**)) and cumulative curves ((**c**) and (**d**)) of Orbitrap MS data for detected molecular formulas in PEfOM. Venn diagram of molecular formulas in the PEfOM and NOM samples under (+)ESI and (−)ESI modes ((**e**) and (**f**)). The red dotted line indicates 95% of the total number of compounds detected ((a) and (b)). Compounds are color coded by the percentage of samples in which they occurred ((c) and (d)).

characteristics, indicating relative homogeneity across its composition. For PEfOM, however, a much larger sample set is required to capture its generalizable features. This directly demonstrates that the total molecular diversity and spatial heterogeneity of PEfOM across China. The molecular diversity of PEfOM and NOM were also reflected in the numbers of molecules. The total number of unique molecular formulas in PEfOM samples surpassed that of NOM samples under both (−)ESI mode (28,442 vs. 7476) and (+)ESI mode (8952 vs. 4384) (Fig. 2(e) and (f)) (Chen et al., 2023; Liu et al., 2023). Notably, the overlap in molecular formulas shared between PEfOM and NOM was only 10% of the overall PEfOM molecular population in (+)ESI mode and 1% in

(−)ESI mode (Fig. 2(e) and (f)). This difference in overlap indicated that the molecular composition of PEfOM was more diverse than that of NOM.

PEfOM exhibited distinct molecular characteristics compared with NOM in both (+)ESI and (−)ESI modes. The MWs of PEfOM samples were mainly between 200 and 380 Da, and were notably lower than those of NOM (340–400 Da) (Fig. 3(a) and (b)). The weighted average DBE values (3–7) and weighted average AI_{mod} values (0–0.3) of PEfOM were also lower than those of NOM (DBE > 7 and AI_{mod} > 0.2) in both (+)ESI and (−)ESI modes (Fig. 3(c) and 3(f)). These results showed that molecules in PEfOM had lower unsaturation and aromaticity than those in NOM, and agreed with the lower $SUVA_{254}$ values (Section 3.1). Given that stable aromatic rings are highly resistant to microbial attack (Mermer et al., 2021), the low aromaticity of PEfOM indicates a lack of such stable aromatic ring structures in its molecules. It implies that a larger proportion of its molecules are readily available for microbial metabolism. Consequently, the molecular composition of PEfOM suggested it had higher biodegradability potential than NOM after discharge into receiving water bodies. The weighted average N/C (0.02–0.12) of PEfOM under (+)ESI mode and the weighted average S/C (0.03–0.08) of PEfOM under (−)ESI mode were both much higher than those in NOM (both approximately zero) (Fig. 3(g) and (h)). The nitrogen- and sulfur-containing components in the PEfOM would form from heteroatom compounds contained in the petrochemical raw materials. High contents of sulfur and nitrogen can promote the adsorption of PEfOM onto sediments or organic colloids via electrostatic interactions and potentially prolong the persistence of PEfOM in aquatic systems (Mallappa et al., 2025). Notably, the weighted average O/C values (0.1–0.4) and NOSC (−1 to 0.6) of PEfOM were greatly lower than those of NOM in (+)ESI mode, whereas the values were relatively similar in PEfOM and NOM in (−)ESI mode (Fig. 3(i) and 3(l)). These results indicated that the basic compounds in PEfOM detected in (+)ESI mode were more susceptible to oxidation in natural water. The difference in thermodynamic reactivity between PEfOM and NOM was evaluated using the ΔG_0 values. In (+)ESI mode, the ΔG_0 values of PEfOM (76–90) covered a higher range than those of NOM (64–76) (Fig. 3(m) and (n)). In (−)ESI mode, this difference was less pronounced. Thus, the molecular chemical stability of basic PEfOM detected in (+)ESI mode was lower than that of NOM under aerobic conditions. These results agreed with the O/C and NOSC results.

The distribution of molecular features showed that PEfOM exhibited unique molecular characteristics with a relatively high heteroatom content, low degree of oxidation, low unsaturation, low MW, and high chemical reactivity compared with NOM. These differences will make PEfOM more liable to biological and chemical transformation in receiving water bodies than NOM. Consequently, the molecular composition and toxicity of PEfOM may be altered after discharge (Niu et al., 2024).

The differences between PEfOM and NOM were further compared in terms of the species present in the samples. For the convenience of discussion, the modified aromaticity index (AI_{mod}) and H/C were used to divide the Van Krevelen diagrams of PEfOM into the following five regions as shown in Fig. S4: saturated compounds (Satu; H/C > 2.0), aliphatic compounds (Aliph; 1.5 < H/C < 2.0), highly unsaturated and phenolic compounds (Huph; AI_{mod}< 0.5, H/C < 1.5), polyphenols (Plph; 0.5 < AI_{mod} < 0.66), and condensed polycyclic aromatics (Cpla; AI_{mod} > 0.66) (Koch and Dittmar 2015). In both (+)ESI and (−)ESI modes, NOM samples were predominantly contained Huph compounds with high stability (Table S2 and Fig. S4) (D'Andrilli et al., 2015). The sum of RIs for Huph and Aliph compounds in the PEfOM samples was >50% (Fig. 4(a) and (b)). These compounds are typically associated with biological metabolites and oxidation intermediates generated in wastewater treatment processes (Kou et al., 2023a). Alhough Huph and Aliph were the dominant species in the PEfOM samples, the RI distributions for these compounds had relatively large standard deviations ((−)ESI mode: ±10.87% for Huph and ±7.00% for Aliph, (+)ESI mode: ±9.63% for Huph and ±14.24% for Aliph). Wide distributions in compositions were also observed for heteroatom compounds in the 21 PEfOM samples. Under (−)ESI mode, the dominant compounds were CHO (33.11% ± 13.23%) and CHOS (49.76% ± 14.44%) (Fig. 4(c)). CHON compounds were dominant (54.36% ± 17.21%) in (+)ESI mode (Fig. 4(d)). The large standard deviations in the contents of major species indicated there were significant variations in molecular compositions among the different PEfOM samples. This variation is in contrast to the uniform molecular compositions of DOM from natural water bodies over large spatial scales (Chen et al., 2023). Consequently, PEfOM not only differs from NOM in terms of overall molecular attributes but also exhibits significant differences in its molecular composition among different PEfOM samples. Therefore, there is a high degree of heterogeneity in PEfOM.

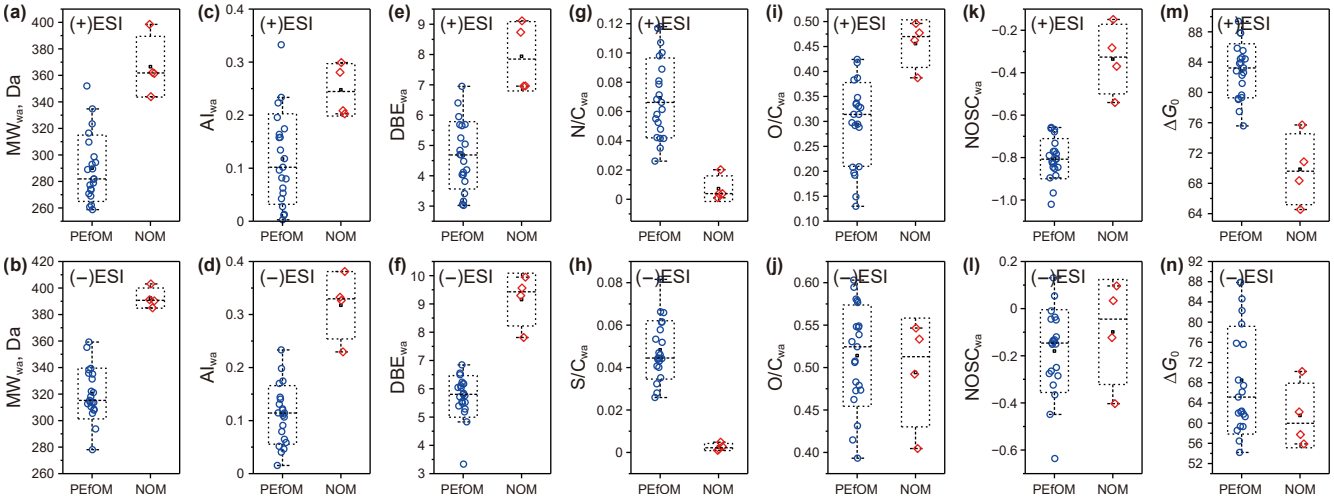


Fig. 3. Molecular attributes (MW_{wa} ((**a**) and (**b**)), AI_{wa} ((**c**) and (**d**)), DBE_{wa} ((**e**) and (**f**)), N/C_{wa} ((**g**) and (**h**)), O/C_{wa} ((**i**) and (**j**)), $NOSC_{wa}$ ((**k**) and (**l**)) and ΔG_0 ((**m**) and (**n**)) of DOM in petrochemical effluent organic matter (PEfOM) and natural organic matter (NOM) in (+)ESI and (−)ESI modes.

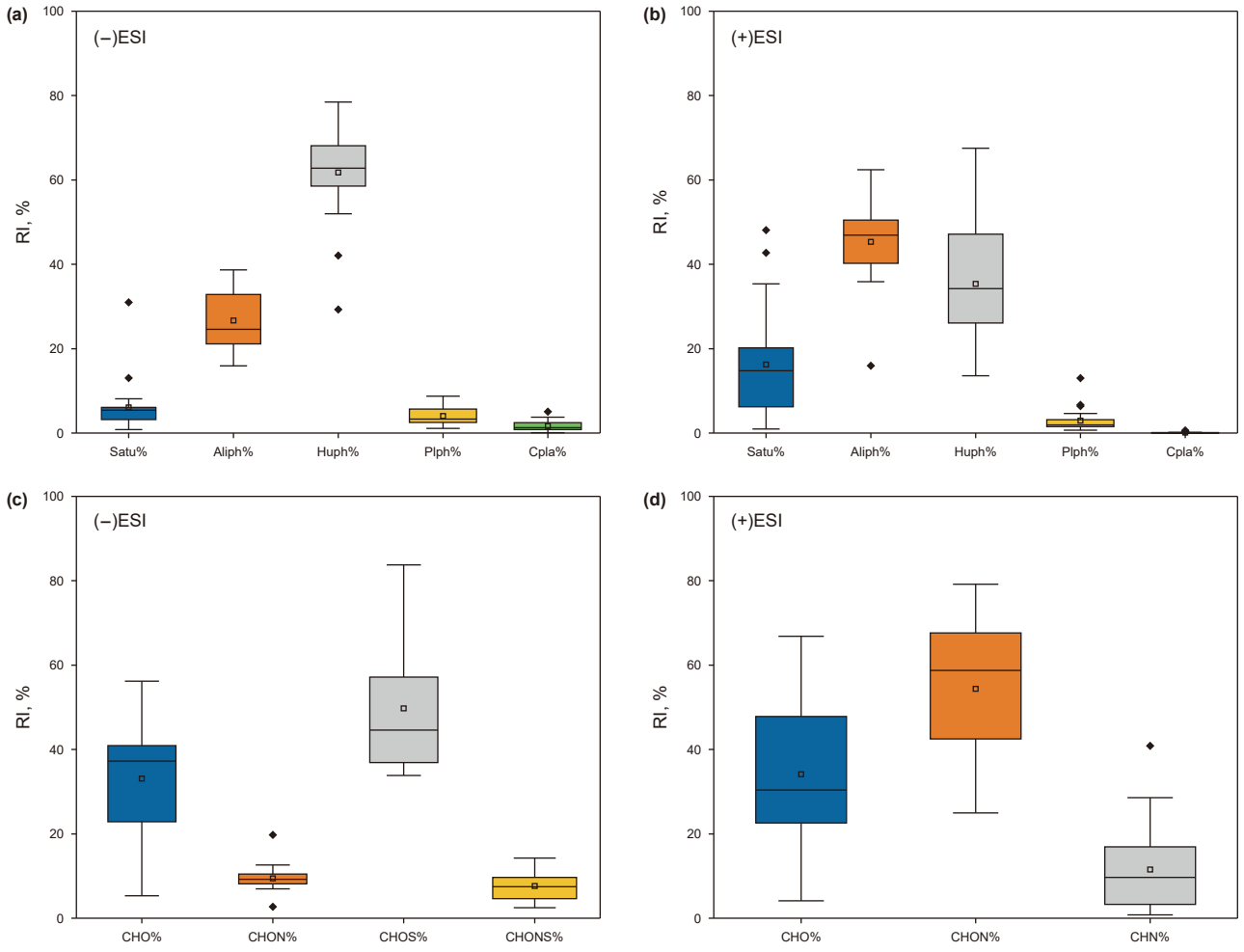


Fig. 4. RIs of five components ((**a**) and (**b**)) and heteroatom compounds ((**c**) and (**d**)) in the PEfOM samples obtained in (+)ESI and (−)ESI modes.

3.3. Identification of the heterogeneity among PEfOM samples

The heterogeneity among PEfOM samples made it challenging to summarize how the overall compositional patterns differed from those of NOM. Therefore, cluster analysis was conducted to extract the common molecular characteristics of PEfOM samples. Considering the significant differences in compositional patterns among PEfOM samples, the JSd method was chosen for cluster analysis. The calculation results exhibited a broad distribution ranging from 0 to 0.2 in both (+)ESI and (−)ESI modes (Fig. S5). Combination of the JSd values and cluster analysis was used to group 20 of the 21 PEfOM samples into three distinct clusters (+CL1, +CL2, and +CL3) for (+)ESI mode, with silhouette coefficients ranging from 0.4 to 0.8 (Fig. 5(a)). Clusters + CL1, +CL2, and +CL3 contained eight, four, and eight samples, respectively (Fig. 5(b)). For (−)ESI mode, 20 of the 21 PEfOM samples were grouped into cluster −CL1 (five samples) and cluster −CL2 (15 samples), with silhouette coefficients ranging from 0.1 to 0.6 (Fig. 5(c) and (d)). Sample #11 in (+)ESI mode and sample #01 in (−)ESI mode had negative silhouette coefficients during cluster analysis, so were not contained in these clusters and were excluded from further discussion. Because the number of clusters under (+)ESI mode was larger than that under (−)ESI mode, we deduced that the basic molecules detected by (+)ESI mode were more heterogeneous than the acidic molecules detected by (−)ESI mode in the PEfOM samples. Since each cluster contains samples derived from crude oils with a wide spectrum of density, sulfur content, and acid value, samples within a cluster showed no clear link to the type of crude oil processed by the plant (Table S3). While the prevalence of treatment-specific molecular signatures (e.g., microbial metabolites and recalcitrant additives) suggests that the great effects of crude oil processing and wastewater treatments on the composition of PEfOM. Despite variations in crude oil sources, the biological and physicochemical units within wastewater treatment plants (WWTPs) act as powerful reactors that dramatically reshape the molecular composition of dissolved organic matter (DOM) (Wen et al., 2023), such as generating soluble microbial products and extracellular polymeric substances (Chen et al., 2024). Furthermore, the use of specific chemical additives (e.g. scale inhibitors, coagulants, and surfactants) introduced during both refining and water treatment stages, contributes directly to the PEfOM pool. Therefore, the combined effects of microbial metabolic activities and process-specific chemical/operational conditions override the initial variations in crude oil composition, becoming the primary drivers of the observed spatial heterogeneity in PEfOM.

To precisely describe the key features of different PEfOM samples, the IndVal was used to identify the characteristic molecular sets associated with each cluster. Compounds with significant IndVal results ($P < 0.05$) are shown in Fig. 6(a) and (e). In (+)ESI mode, the characteristic molecular set of cluster +CL1 was primarily composed osf nitrogen-poor (N/C = 0–0.05) but oxygen-

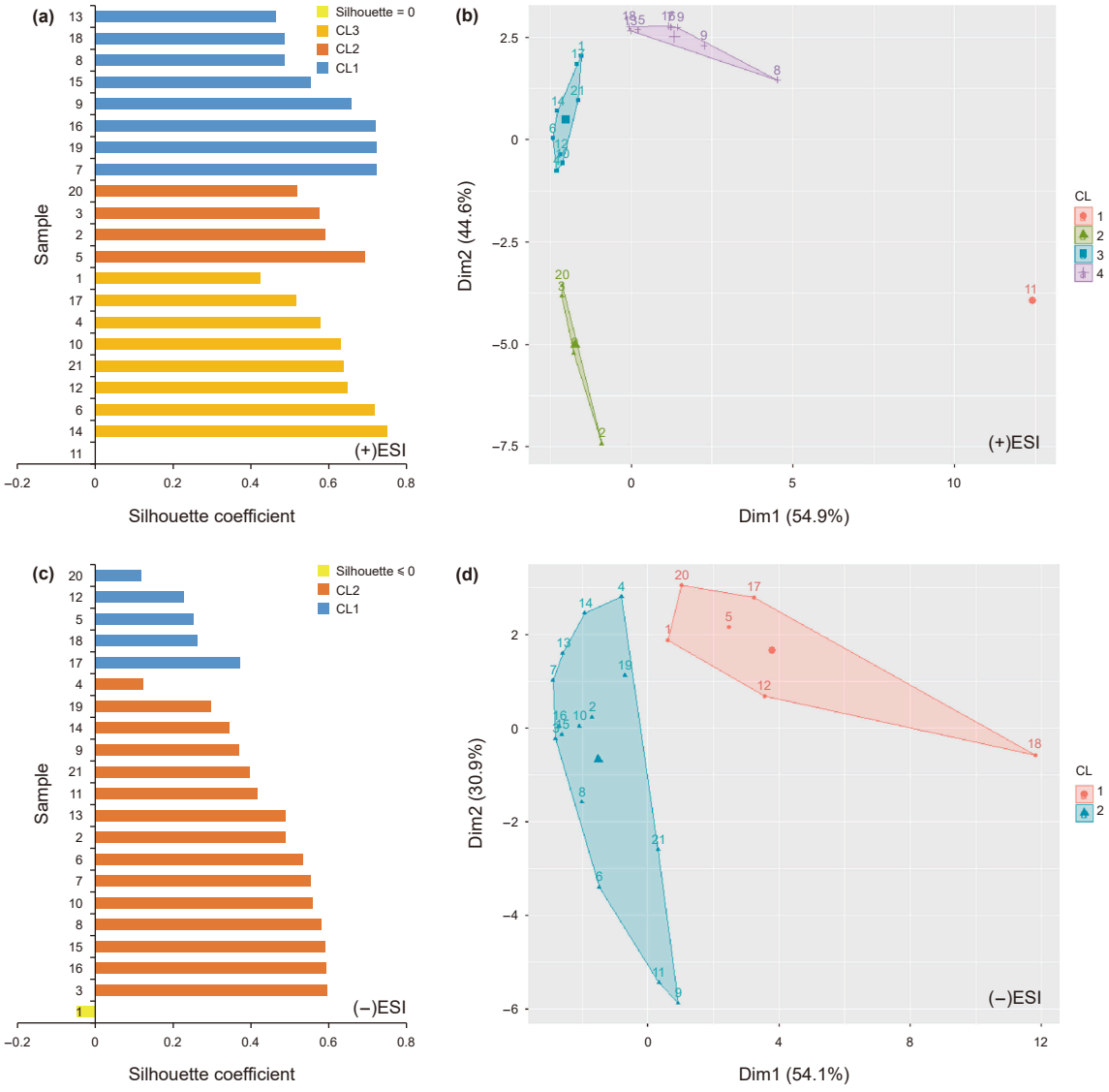


Fig. 5. Hierarchical clustering and silhouette coefficients for all PEfOM samples in (+)ESI mode ((**a**) and (**b**)) and (−)ESI mode ((**c**) and (**d**)). Sample 11 in (+)ESI mode and sample #01 in (−)ESI mode had negative silhouette coefficients and were excluded from further analysis.

rich (O/C > 0.5) components with low degrees of unsaturation (DBE = 0–8) (Fig. S6). These compounds likely represented oxidation intermediates from biological wastewater treatment processes (He et al., 2023). The high oxygen contents suggested high bioavailability, which may accelerate microbial degradation in receiving water bodies (Xie et al., 2024). The characteristic molecular set of cluster +CL2 was enriched in CHON compounds with high unsaturation (DBE = 4–14) and low O/C (<0.2). Some unique CHN compounds were also found in the characteristic molecular set of cluster +CL2 (Fig. S6). These compounds are usually refractory heterocyclic compounds (e.g., quinoline derivatives) from point-source petrochemical wastewater, which are resistant to degradation in conventional biological wastewater treatment units. Notably, those characteristic nitrogen-containing compounds in cluster + CL2 had higher ΔG_0 values (80–110) than the characteristic compounds in cluster +CL1 (ΔG_0 = 50–80) and cluster +CL3 (ΔG_0 = 60–90) (Fig. 6(f)). These ΔG_0 values showed that the nitrogen-containing compounds in cluster +CL2 were more thermodynamically favorable in aerobic environments than the compounds in clusters +CL1 and +CL3. The transformation products may pose potential ecological risks because of their persistence and toxicity (He et al., 2023). The characteristic molecular set of cluster +CL3 was associated with a limited number of molecular formulas characterized by O/C of 0.2–0.5 and DBE of 4–14 (Fig. S6). Some compounds containing one nitrogen atom (e.g., $-NO_2$) may be associated with byproducts formed during biological wastewater treatment. Thus, the characteristic molecular sets of PEfOM were mainly oxygen-rich or N-containing compounds under (+)ESI mode, and were significantly influenced by the transformation of DOM in biological wastewater treatment units. Nitrogen-containing compounds originating from point-source wastewater also affected the distribution of basic nitrogen-containing compounds in PEfOM, which showed high reactivity after discharge to natural water bodies.

In (−)ESI mode, the characteristic molecular set of cluster −CL1 was dominated by oxygen-poor (O/C < 0.5) molecules with high MWs (600–800 Da) and DBE (2–12). Although the overall molecular weight of PEfOM is relatively low, the presence of these specific high-MW, oxygen-poor compounds can be attributed to recalcitrant industrial additives and their derivatives, such as

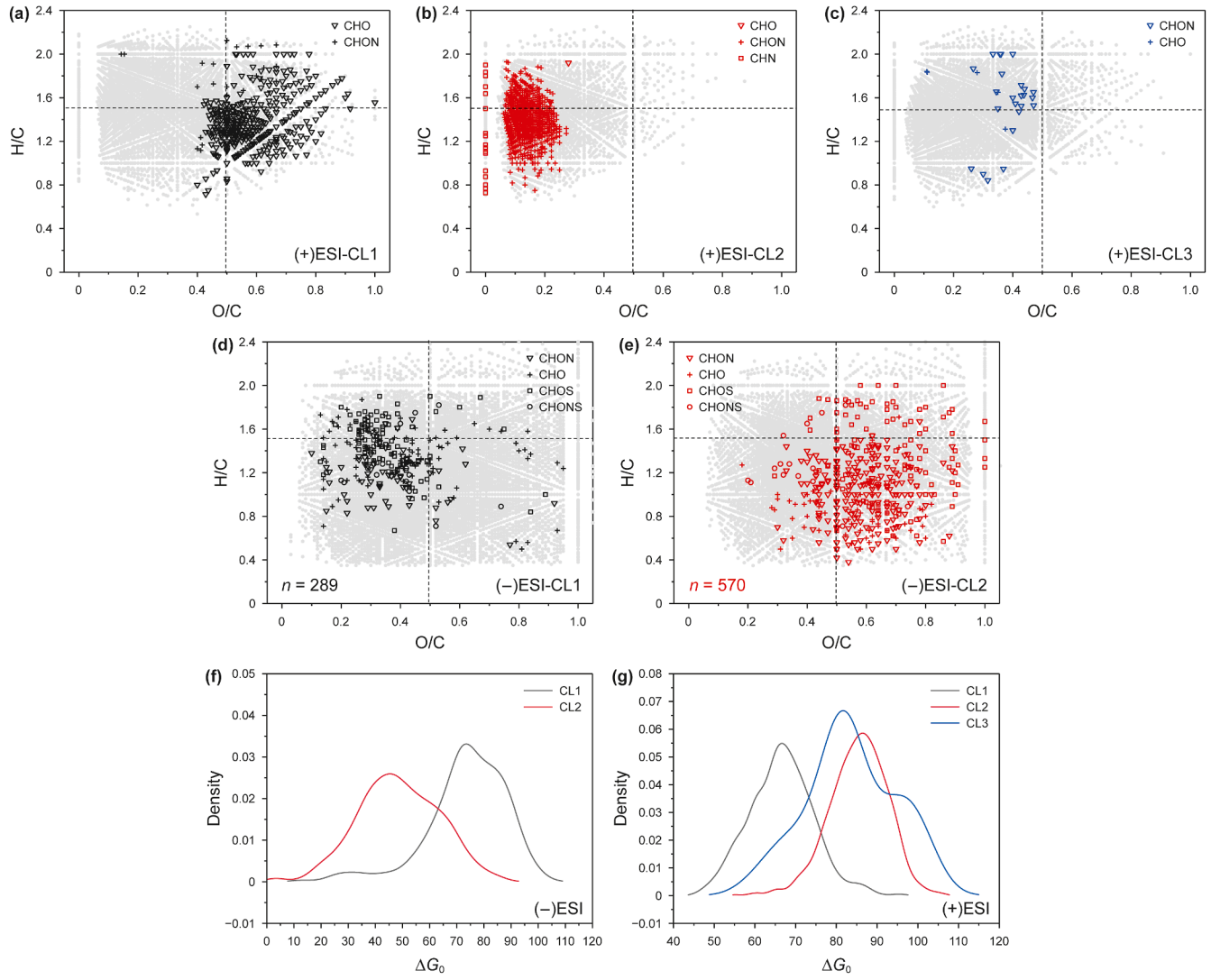


Fig. 6. Molecular sets that are significantly associated with each cluster under (+)ESI mode ((**a**)–(**c**)) and (−)ESI mode ((**d**)–(**e**)). Formulas are plotted in the van Krevelen space with different shapes representing different heteroatoms. Background gray dots are molecular formulas in each molecular set. The dotted lines indicate the boundary between unsaturated and highly unsaturated molecular formulas (H/C = 1.5) and high and low oxygen contents (O/C = 0.5). Distribution of ΔG_0 for molecular sets obtained in (+)ESI mode (**f**) and (−)ESI mode (**g**).

sulfonated surfactants or polymeric scale inhibitors used in refinery processes (Bisatto et al., 2022). Additionally, partially degraded residues of high-MW hydrocarbons from crude oil or microbial by-products generated during wastewater treatment could also contribute to this fraction. The ΔG_0 values of the characteristic molecular set for cluster −CL1 also covered a wide range (60–100), which indicated the compounds were photochemically reactive in an aerobic environment. The characteristic molecular set of cluster −CL2 contained mainly oxygen-rich molecules (O/C = 0.4–1) with low ΔG_0 values (20–80) (Figs. S7 and 6(g)). High O/C values are associated with high hydrophilicity and molecular mobility in water (Davoudi et al., 2025). In summary, in addition to many metabolic intermediates from biological units, PEfOM was also dominated by recalcitrant basic nitrogen-containing compounds and acidic sulfur-containing compounds. These components exhibited relatively high reactivity under aerobic conditions, and their potential for further transformation in aquatic environments is concerning.

3.4. Representative compounds and their biotoxicities in PEfOM

To evaluate the environmental risks of PEfOM in detail, it was necessary to identify the representative basic nitrogen-containing compounds and acidic sulfur-containing compounds within the clusters that were specific to each cluster. To achieve this, the IndVal was applied to prioritize molecules that were highly specific to each cluster (IndVal >0.7, $P < 0.05$) (Table S4). A high IndVal signifies that a molecule is not only abundant but also has a high fidelity to a particular cluster, meaning it is found in most samples of that cluster and is rare or absent in others. This prioritization ensured relevance of the molecules to source characterization and fate prediction. $O_{7–12}$ species in cluster +CL1 had the highest IndVal of 1, which showed that these species were only detected in samples in this cluster and that they fully represented the molecular characteristics of the cluster. The representative $O_{7–12}$ species were also consistent with the overall oxygen-rich and highly hydrophilic characteristics of cluster +CL1 (Section 3.3). For

cluster +CL2, $C_{29–33}HON$ compounds with high aromaticity (AI_{mod} > 0.6, DBE = 10–14) had the highest IndVal (0.94–1), and were selected as representative nitrogen-containing heterocyclic compounds in this cluster. Cluster +CL3 contained a mixture of CHON and CHO compounds, with $C_{18}H_{33}N_1O_5$ having the highest IndVal (0.70). For cluster −CL1, $O_{9–13}S_1$ (e.g., $C_{35}H_{48}O_{13}S_1$) were selected as representative sulfur-containing compounds (IndVal = 0.86). These compounds possibly contained one sulfonate group, which would enhance their solubility and mobility in aquatic systems. However, the high MWs of these compounds may hinder their removal by membrane-based technologies (He et al., 2023). The representative compounds in cluster −CL2 were CHONS compounds with carboxyl and sulfonate functional groups (e.g., $C_9H_{17}N_1O_5S_1$, IndVal = 0.97). These polar molecules could act as chelating agents to mobilize heavy metals in water and exacerbate co-contaminant risks (Li et al., 2025).

On the basis of the IndVal and RI results, three representative compounds ($C_{17}H_{23}N_1O_2$ from cluster + CL2, $C_{14}H_{20}O_5S_1$ from cluster −CL1, and $C_9H_{17}N_1O_5S_1$ from cluster −CL2) were selected to obtain the annotated chemical structures through Orbitrap MS/MS analysis (Fig. S8). The MS/MS results for $C_{17}H_{23}N_1O_2$ (*m/z* 274.180, IndVal = 0.77) showed a characteristic ion peak for $C_8H_9N^+$ (Fig. 7(a)), which was formed by loss of the fragments H_2O, CH_2O_2, and CH_4. These fragmentation patterns suggested that quinolinium groups (Level 2 identification) were present in the compounds in PEfOM in cluster +CL2. This result agreed with a previous study where quinolinium compounds were detected in PEfOM in Jilin Province, China (Li et al., 2024a). Quinolinium compounds are commonly found in petrochemical raw materials (e.g., heterocyclic aromatic hydrocarbons from coal tar and crude oil) and catalytic cracking by-products (Srivani and Sastry, 2009), but are extremely rare in NOM. For $C_9H_{17}N_1O_5S$ (*m/z* 250.076, IndVal = 0.97) from cluster −CL2, a characteristic ion peak was observed for HO_3S^-, which was generated through loss of the fragments CH_4, CO, H_2O, CO_2, CH_4O, CH_2O, and $C_5H_9NO_2$ (Fig. 7(b)), indicating the presence of amino and sulfonic acid groups within the molecule in this cluster. Polar functional groups (–NH_2 and –SO_3H) may be formed from the sulfur/nitrogen-containing compounds in crude oil (e.g., thiols, sulfides, thiophenes, pyridine and quinoline) through oxidation/hydrotreating reactions (Shi and Wu, 2021). These compounds are also rarely detected in NOM, but their strong hydrophilicity (O/C = 0.5–1) may promote chelation with heavy metals such as Cu^{2+} and Pb^{3+}, which will increase the risk of complex pollution after emission (Kumar and Chang, 2025). The characteristic ion peak for $C_{14}H_{20}O_5S_1$ (*m/z* 299.096, IndVal = 0.76) was also HO_3S^-, which was generated through loss of the fragments $C_6H_6O_3S$, CH_4, C_2H_4, CO, CH_4O, CH_2O, and H_2O (Fig. 7(c)). We hypothesized that a benzenesulfonic acid group (Level 3 identification), which is a key structure in surfactants, was present in $C_{14}H_{20}O_5S_1$. Sulfonated surfactants are usually produced from detergents and are frequently detected in municipal wastewater. In petrochemical wastewater, sulfonates mainly originate from refinery additives or scale inhibitors, both of which are used to improve the quality of petrochemical products (Ho et al., 2023). Sulfonates in PEfOM reportedly show strong adsorption and persistence (Deng et al., 2022). The functional groups (quinoline, nitro, and sulfonate groups) in the characteristic compounds of PEfOM further indicated that their sources and formation processes were distinct from those of compounds in NOM. The complexity and diversity of PEfOM samples can be attributed to unique sources and formation pathways of compounds related to the petrochemical industry.

Acute biotoxicities (median lethal concentration (LC_{50}) or half maximal effective concentration (EC_{50})) of representative compounds were estimated using the Ecological Structure Activity Relationships model for fish (96 h LC_{50}), green algae (96 h EC_{50}), and *Daphnia* (48 h LC_{50}) (Table 1). The LC_{50} or EC_{50} values were used to categorize the biotoxicities of the compounds as follows: highly toxic (0–1 mg/L), toxic (1–10 mg/L), harmful (10–100 mg/L), and non-toxic (>100 mg/L). The compound $C_{17}H_{23}N_1O_2$ had two possible isomers. One isomer had alkyl groups on both the benzene and pyrrole rings and was highly toxic to fish (LC_{50} = 0.064 mg/L) and green algae (EC_{50} = 0.36 mg/L), and toxic to *Daphnia* (LC_{50} = 1.59 mg/L). The other isomer only had substituents on the pyrrole ring and was harmful to fish (LC_{50} = 33 mg/L) and toxic to green algae (EC_{50} = 4.1 mg/L) and *Daphnia* (LC_{50} = 3.1 mg/L). By contrast, the compounds $C_{14}H_{20}O_5S_1$ and $C_9H_{17}N_1O_5S_1$ were non-toxic towards the same aquatic organisms, with LC_{50} or EC_{50} values exceeding 1000 mg/L. Although the Ecological Structure Activity Relationships model provides a rapid screening tool for toxicity prediction, its reliance on quantitative structure–activity relationships may introduce uncertainties for complex industrial compounds with limited experimental data (Achar et al., 2024). For example, synergistic effects of co-existing pollutants or site-specific environmental conditions (e.g., pH and DOC) are not accounted for in this model. The potential toxicity risks of nitrogen-containing compounds in PEfOM require continued attention and further laboratory bioassays (e.g., zebrafish embryo tests or algal growth inhibition experiments). Furthermore, for effluents characterized by nitrogen-containing heterocyclic compounds, such as those in cluster +CL2, the transformation behavior needs to be explored after discharge into natural water bodies.

3.5. Environmental implications

The identification of spatial heterogeneity and characteristic compounds of PEfOM provides critical insight for source tracking and targeted treatment of petrochemical effluents. For instance, the quinolinium derivative ($C_{17}H_{23}N_1O_2$) was predominantly detected in samples belonging to +CL2, with markedly lower relative intensities in samples from +CL1 and +CL3. Similarly, the sulfonated compounds $C_{14}H_{20}O_5S_1$ and $C_9H_{17}N_1O_5S_1$ were overwhelmingly characteristic of −CL1 and −CL2, respectively. This non-uniform distribution provides direct molecular evidence supporting the spatial heterogeneity of PEfOM, demonstrating that different petrochemical plants discharge effluents with distinct chemical fingerprints. The high sulfur/nitrogen heteroatom content and industry-specific structures (e.g., quinolinium) in PEfOM can serve as molecular markers for tracing petrochemical pollution sources. For instance, the characteristic sulfonates in PEfOM may be related to differences in the crude oil sulfur content and processing technologies (e.g. catalytic cracking vs. hydrodesulfurization) among different petrochemical plants. For quinolinium and other difficult-to-degrade substances, advanced oxidation processes (e.g. using ozone or the electro-Fenton process) should be integrated into existing processes (e.g. activated sludge) to improve removal rates (El-Naas and Al-Khalid, 2017). This study identifies quinolinium and sulfonate compounds as characteristic molecular markers for PEfOM. It is important to note that the non-targeted Orbitrap-MS approach used here is not fully quantitative. The environmental significance of compounds based on their concentrations requires further validation using targeted analytical methods with appropriate internal standards. Besides, the Orbitrap-MS used in this study is suitable for detecting compounds with high mass spectrometry responses. To provide more comprehensive results for petrochemical effluents, robust extraction and quantification techniques, such as high-performance liquid chromatography-Orbitrap MS, are needed to improve the

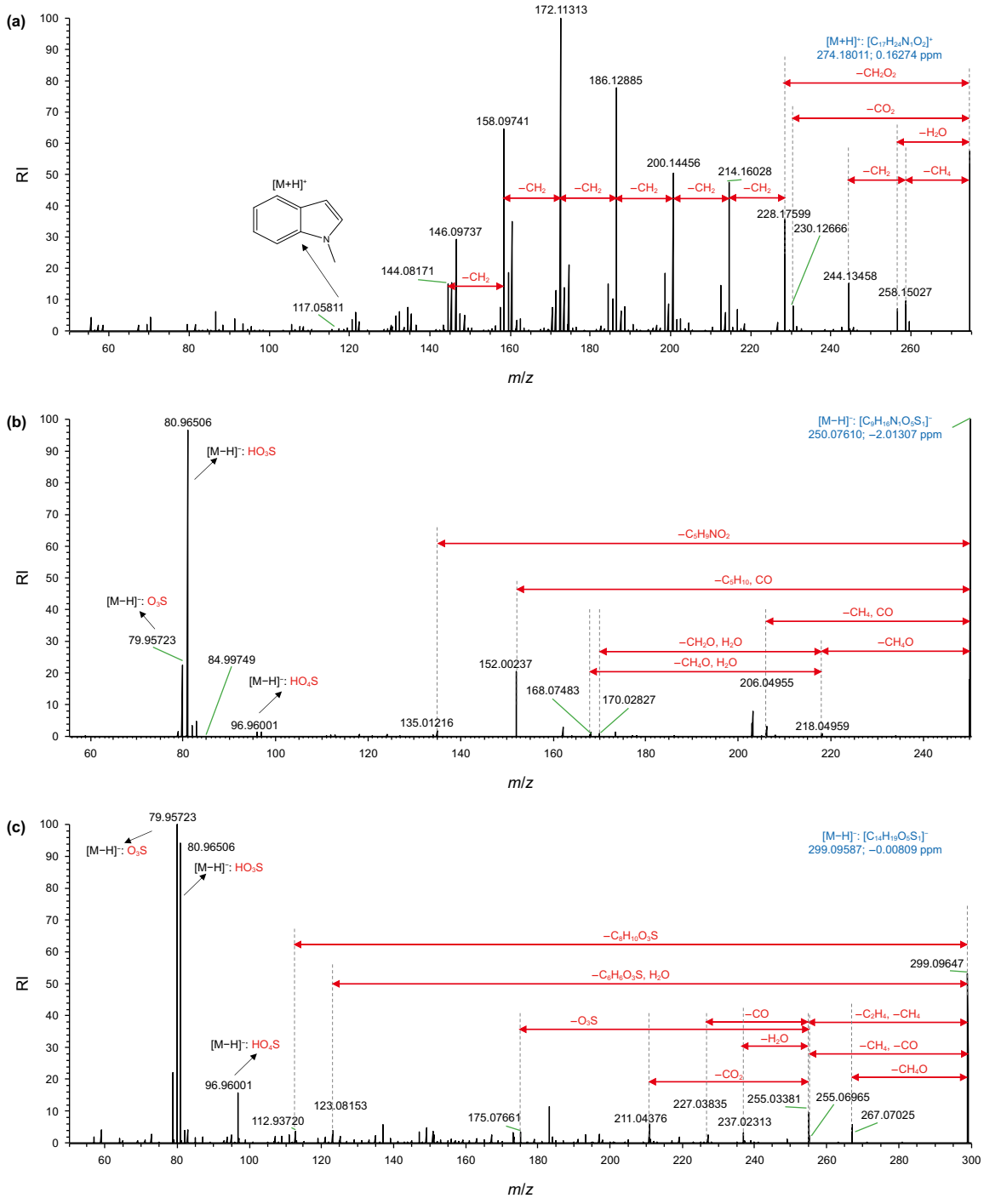
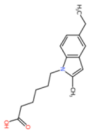
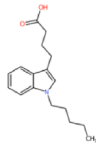
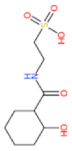
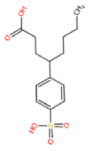
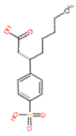


Fig. 7. Tandem mass spectrometry results for representative compounds at *m/z* 274.180 under (+)ESI mode (**a**) and at *m/z* 250.076 and 299.096 under (−)ESI mode ((**b**) and (**c**)).

detection of compounds in PEfOM with low mass spectrometry responses (Liang et al., 2024).

Although the PEfOM samples were highly heterogeneous, their overall differences compared with NOM were greater than this heterogeneity. Thus, it is necessary to incorporate molecular-level DOM monitoring into existing emission standards for the petrochemical industry to facilitate comprehensive assessment of PEfOM emissions. The Orbitrap-MS used in this study is suitable for detecting compounds with high mass spectrometry responses. To provide more comprehensive results for petrochemical effluents, alternative analytical methods, such as high-performance liquid chromatography-Orbitrap MS, are needed to improve the detection of compounds in PEfOM with low mass spectrometry responses (Liang et al., 2024). Moreover, some nitrogen-containing heterocyclic compounds in PEfOM displayed particularly high biotoxicity and reactivity compared with NOM, which may lead photochemical and microbial degradation processes that produce new secondary pollutants in the receiving aquatic environment (An et al., 2024; Xu et al., 2024). Therefore, photo-transformation and microbial transformation of PEfOM and

Table 1
Estimated toxicities of representative compounds.

		$C_{17}H_{23}N_1O_2$ m/z = 274.180 Da		$C_9H_{17}N_1O_5S_1$ m/z = 250.076 Da		$C_{14}H_{20}O_5S_1$ m/z = 299.096 Da	
Structure		Structure-1	Structure-2	Structure -1	Structure -2	Structure -1	Structure-2
							
Biotoxicity, mg/L	Fish (96 h) LC_{50}	0.064	33.0	3.2×10^6	9.8×10^4	6.8×10^3	
	Daphnid (48 h) LC_{50}	1.59	4.1	6.3×10^6	7.8×10^3	3.7×10^3	
	Green algae (96 h) EC_{50}	0.36	3.1	7.0×10^6	1.4×10^4	2.3×10^3	

the potential effects on ecological and human health should be further explored to increase our understanding of the environmental behavior of petrochemical effluents.

4. Conclusions

In this study, (+)ESI and (−)ESI modes in Orbitrap MS were combined to overcome the limitations of single ionization techniques and comprehensively capture information for both acidic and basic components in PEfOM on a large scale. Compared with NOM, PEfOM had relatively low aromaticity, a low MW, and a high heteroatom content, and contained Aliph and Huph as the dominant components. Significant spatial heterogeneity in the molecular composition of 21 PEfOM samples was observed across different plants. Three clusters of molecules were obtained for PEfOM in (+)ESI mode and two clusters in (−)ESI mode. These clusters showed significant differences in heteroatom-containing species, unsaturation, and aromaticity. Nitrogen-rich CHON compounds with high unsaturation in cluster +CL2, and oxygen-poor compounds with high MWs in cluster −CL1 exhibited the greatest reactivity in aquatic environments under aerobic conditions. $C_{17}H_{23}N_1O_2$, $C_{14}H_{20}O_5S_1$, and $C_9H_{17}N_1O_5S_1$ were identified as representative compounds in PEfOM. Notably, $C_{17}H_{23}N_1O_2$, which contains a quinoline structure, was confirmed to have high acute biotoxicity. There is a need for focused attention on the accumulation risks of these compounds in the environment. These results highlight the necessity for targeted monitoring and management strategies to mitigate potential ecological effects of PEfOM.

CRediT authorship contribution statement

Yue Kou: Writing – original draft, Formal analysis. **Huang-Fan Ye:** Project administration. **Rui Zhang:** Investigation. **Wen-Yi Lu:** Methodology. **Ze-Lin Niu:** Investigation. **Yi-Fei Zhu:** Investigation. **Zhuo-Yu Li:** Writing – review & editing, Funding acquisition. **Chen He:** Methodology. **Hai-Feng Wang:** Funding acquisition. **Qing-Hong Wang:** Validation. **Quan Shi:** Validation. **Chun-Mao Chen:** Supervision.

Declaration of competing interest

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

Acknowledgements

This study was supported in part by the National Key Research and Development Program (Grant No. 2024YFF0618501), National Natural Science Foundation of China (Grant No. 21776220), Science Foundation of China University of Petroleum, Beijing (Grant No. 2462024YJRC004) and the fundamental research funds of the National Institute of Metrology (Grant No. AKYZZ2551).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petsci.2026.03.021.

References

Achar, J., Firman, J.W., Cronin, M.T.D., Öberg, G., 2024. A framework for categorizing sources of uncertainty in in silico toxicology methods: considerations for chemical toxicity predictions. Regul. Toxicol. Pharmacol. 154, 105737. https://doi.org/10.1016/j.yrtph.2024.105737.

An, X., Li, N., Zhang, L., Xu, Z., Zhang, S., Zhang, Q., 2024. New insights into the typical nitrogen-containing heterocyclic compound-quinoline degradation and detoxification by microbial consortium: Integrated pathways, meta-transcriptomic analysis and toxicological evaluation. J. Hazard Mater. 465, 133158. https://doi.org/10.1016/j.jhazmat.2023.133158.

Bao, M., Fang, F., Luo, X., Li, Q., 2024. Transformation characteristics of dissolved organic matter in landfill leachate nanofiltrated concentrate during ozonation-based advanced oxidation systems. J. Environ. Manag. 12 (2), 112076. https://doi.org/10.1016/j.jece.2024.112076.

Bisatto, R., Picoli, V.M., Petzhold, C.L., 2022. Evaluation of different polymeric scale inhibitors for oilfield application. J. Pet. Sci. Eng. 213, 110331. https://doi.org/10.1016/j.petrol.2022.110331.

Cai, R., Jiao, N., 2023. Recalcitrant dissolved organic matter and its major production and removal processes in the ocean. Deep-sea Res. Pt. I 191, 103922. https://doi.org/10.1016/j.dsr.2022.103922.

Chen, L.M., Erol, O., Choi, Y.H., Pronk, M., van Loosdrecht, M., Lin, Y., 2024. The water-soluble fraction of extracellular polymeric substances from a resource recovery demonstration plant: characterization and potential application as an adhesive. Sec. Microbiotechnology 15. https://doi.org/10.3389/fmicb.2024.1331120.

Chen, G., Qian, C., Gong, B., Du, M., Sun, R.Z., Chen, J.J., Yu, H.Q., 2023. Unraveling heterogeneity of dissolved organic matter in highly connected natural water bodies at molecular level. Water Res. 246, 120743. https://doi.org/10.1016/j.watres.2023.120743.

D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G., 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Commun. Mass Spectrom. 29 (24), 2385–2401. https://doi.org/10.1002/rcm.7400.

Davoudi, M., Nezhad, F.K., Rahdar, S., Igwegbe, C.A., 2025. Current applications of functionalized reduced graphene oxide-based semiconductors for photocatalytic removal of pollutants from wastewater: A review. Chem. Pap. 79 (2), 637–654. https://doi.org/10.1007/s11696-024-03816-5.

Deng, W., He, C., Shi, Q., Huang, H., 2022. Separation and characterization of sulfonates in dissolved organic matter from industrial wastewater by solid phase

extraction and high-resolution mass spectrometry. Anal. Bioanal. Chem. 414 (16), 4697–4706. https://doi.org/10.1007/s00216-022-04092-6.
Du, P., Chen, G., Zhang, P., Yang, B., Wang, J., 2023. Photo-transformation of wastewater effluent organic matter reduces the formation potential and toxicity of chlorinated disinfection byproducts. Ecotoxicol. Environ. Saf. 265, 115515. https://doi.org/10.1016/j.ecoenv.2023.115515.
El-Naas, M., Al-Khalid, T., 2017. Organic contaminants in refinery wastewater: Characterization and novel approaches for biotreatment. In: Zoveidavianpoor, M. (Ed.), Recent Insights in Petroleum Science and Engineering. IntechOpen, Rijeka.
Fatkullina, A.F., Zakharova, N.V., Akhmetzyanova, G.S., 1979. Selective chromatographic determination of sulfur-containing compounds in work-area atmosphere. Chem. Technol. Fuels Oils 15 (4), 252–253. https://doi.org/10.1007/BF00721733.
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 (8), 4374–4381. https://doi.org/10.1021/es2041114.
Harjung, A., Schweichhart, J., Rasch, G., Griebler, C., 2023. Large-scale study on groundwater dissolved organic matter reveals a strong heterogeneity and a complex microbial footprint. Sci. Total Environ. 854, 158542. https://doi.org/10.1016/j.scitotenv.2022.158542.
Hawkes, J.A., Dittmar, T., Patriarca, C., Tranvik, L., Bergquist, J., 2016. Evaluation of the orbitrap mass spectrometer for the molecular fingerprinting analysis of natural dissolved organic matter. Anal. Chem. 88 (15), 7698–7704. https://doi.org/10.1021/acs.analchem.6b01624.
He, D., He, C., Li, P., Zhang, X., Shi, Q., Sun, Y., 2019. Optical and molecular signatures of dissolved organic matter reflect anthropogenic influence in a coastal river, Northeast China. J. Environ. Qual. 48 (3), 603–613. https://doi.org/10.2134/jeq2018.09.0330.
He, C., Liu, J., Wang, R., Li, Y., Zheng, Q., Jiao, F., He, C., Shi, Q., Xu, Y., Zhang, R., Thomas, H., Batt, J., Hill, P., Lewis, M., MacIntyre, H., Lu, L., Zhang, Q., Tu, Q., Shi, T., Chen, F., Jiao, N., 2022. Metagenomic evidence for the microbial transformation of carboxyl-rich alicyclic molecules: A long-term macrocosm experiment. Water Res. 216, 118281. https://doi.org/10.1016/j.watres.2022.118281.
He, C., Chen, W., Chen, C., Shi, Q., 2023. Molecular transformation of dissolved organic matter in refinery wastewaters: Characterized by FT-ICR MS coupled with electrospray ionization and atmospheric pressure photoionization. Pet. Sci. 20 (1), 590–599. https://doi.org/10.1016/j.petsci.2022.09.035.
Ho, Q.N., Anam, G.B., Kim, J., Park, S., Lee, T.-U., Jeon, J.-Y., Choi, Y.-Y., Ahn, Y.-H., Lee, B.J., 2023. Fate of sulfate in municipal wastewater treatment plants and its effect on sludge recycling as a fuel source. Sustainability 15 (1), 311. https://doi.org/10.3390/su15010311.
Kellerman, A.M., Dittmar, T., Kothawala, D.N., Tranvik, L.J., 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat. Commun. 5, 3804. https://doi.org/10.1038/ncomms4804.
Kida, M., Merder, J., Fujitake, N., Tanabe, Y., Hayashi, K., Kudoh, S., Dittmar, T., 2023. Determinants of microbial-derived dissolved organic matter diversity in Antarctic Lakes. Environ. Sci. Technol. 57 (13), 5464–5473. https://doi.org/10.1021/acs.est.3c00249.
Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Commun. Mass Spectrom. 20 (5), 926–932. https://doi.org/10.1002/rcm.2386.
Kou, Y., Jiang, J., Yang, B., Sun, H., Wang, L., Wang, Q., El-Din, M.G., Shi, Q., Chen, C., 2023a. Transformation of dissolved organic matter at a full-scale petrochemical wastewater treatment plant. J. Environ. Manag. 329, 117021. https://doi.org/10.1016/j.jenvman.2022.117021.
Kou, Y., Yang, B., Jiang, J., Sun, H., Zhang, R., Li, Z., Wang, Q., Shi, Q., Chen, C., 2023b. Characteristics of dissolved organic matter in point-source wastewaters at a petrochemical plant: Molecular constituents and contributions to the influent of wastewater treatment plant. Environ. Res. 238, 117157. https://doi.org/10.1016/j.envres.2023.117157.
Kumar, A., Chang, D.W., 2025. Active polymers decorated with major acid groups for water treatment: Potentials and challenges. Polymers 17 (1), 29. https://doi.org/10.3390/polym17010029.
Laszakovits, J.R., Somogyi, A., MacKay, A.A., 2020. Chemical alterations of dissolved organic matter by permanganate oxidation. Environ. Sci. Technol. 54 (6), 3256–3266. https://doi.org/10.1021/acs.est.9b06675.
Li, M., Li, Z., Fu, L., Deng, L., Wu, C., 2024a. Molecular-level insights into dissolved organic matter and its variations of the full-scale processes in a typical petrochemical wastewater treatment plant. Water Res. 261, 121990. https://doi.org/10.1016/j.watres.2024.121990.
Li, Y., He, C., Chen, C., Liu, F., Shi, Q., 2024b. Molecular investigation into the transformation of recalcitrant dissolved organic sulfur in refinery sour water during stripping process. Pet. Sci. 21 (3), 2112–2119. https://doi.org/10.1016/j.petsci.2024.01.005.
Li, Y., He, C., Li, Z., Zhang, Y., Wu, B., Shi, Q., 2020. Molecular transformation of dissolved organic matter in refinery wastewater. Water Sci. Technol. 82 (1), 107–119. https://doi.org/10.2166/wst.2020.334.
Li, Y., He, C., Zhang, Y., Shi, Q., 2024c. Online LC-Orbitrap MS method for the rapid molecular characterization of dissolved organic matter. Rapid Commun. Mass Spectrom. 38 (20), e9885. https://doi.org/10.1002/rcm.9885.
Li, Y., Zhang, S., Fu, H., Sun, Y., Tang, S., Xu, J., Li, J., Gong, X., Shi, L., 2025. Immobilization or mobilization of heavy metal(loid)s in lake sediment-water interface: Roles of coupled transformation between iron (oxyhydr)oxides and natural organic matter. Sci. Total Environ. 959, 178302. https://doi.org/10.1016/j.scitotenv.2024.178302.
Liang, L., Li, Y., Peng, C., Ning, L., Yi, G., Wang, W., Yuan, H., Liu, P., 2024. Pre-column derivative HPLC and LC-Orbitrap-MS analysis of monosaccharides and non-polysaccharides in Polygonati Rhizoma. Chromatographia 87 (9), 549–559. https://doi.org/10.1007/s10337-024-04350-y.
Liao, K., You, J., Han, C., Cheng, H., Ren, H., Hu, H., 2024. Dissolved organic nitrogen depresses the expected outcome of wastewater treatment upgrading on effluent eutrophication potential mitigation: molecular mechanistic insight. Water Res. 267, 122535. https://doi.org/10.1016/j.watres.2024.122535.
Lin, J., 1991. Divergence measures based on the Shannon entropy. Ieee T. Inform. Theory 37 (1), 145–151. https://doi.org/10.1109/18.61115.
Liu, Y., Liu, X., Long, Y., Wen, Y., Ma, C., Sun, J., 2023. Variations in dissolved organic matter chemistry on a vertical scale in the eastern Indian Ocean. Water Res. 232, 119674. https://doi.org/10.1016/j.watres.2023.119674.
Mallappa, Chahar, M., Choudhary, N., Yadav, K.K., Qasim, M.T., Zairov, R., Patel, A., Yadav, V.K., Jangir, M., 2025. Recent advances in the synthesis of nitrogen-containing heterocyclic compounds via multicomponent reaction and their emerging biological applications: A review. Iran. Chem. Soc. 22 (1), 1–33. https://doi.org/10.1007/s13738-024-03142-3.
Mansour, M.S.M., Abdel-shafy, H.I., Ibrahim, A.M., 2024. Petroleum wastewater: Environmental protection, treatment, and safe reuse: An overview. J. Environ. Manag. 351, 119827. https://doi.org/10.1016/j.jenvman.2023.119827.
Mermer, A., Keles, T., Sirin, Y., 2021. Recent studies of nitrogen containing heterocyclic compounds as novel antiviral agents: A review. Bioorg. Chem. 114, 105076. https://doi.org/10.1016/j.bioorg.2021.105076.
Ministry of Ecology and Environment of the People's Republic of China, 2015. Emission Standard of Pollutants for Petroleum Chemistry Industry (GB 31571-2015).
Murphy, K.R., Stedmon, C.A., Wenig, P., Bro, R., 2014. OpenFluor– an online spectral library of auto-fluorescence by organic compounds in the environment. Anal. Methods 6 (3), 658–661. https://doi.org/10.1039/C3AY41935E.
Nguyen, N.T., Kusakabe, T., Takaoka, M., 2024. Characterization and differentiation of dissolved organic matter in leachate derived from an old Japanese landfill site through Orbitrap mass spectrometry. J. Mater. Cycles Waste 26 (4), 2138–2151. https://doi.org/10.1007/s10163-024-01952-4.
Niu, Y., Yan, Y., Dong, J., Yue, K., Duan, X., Hu, D., Li, J., Peng, L., 2024. Evidence for sustainably reducing secondary pollutants in a typical industrial city in China: Co-benefit from controlling sources with high reduction potential beyond industrial process. J. Hazard Mater. 478, 135556. https://doi.org/10.1016/j.jhazmat.2024.135556.
Phungsai, P., Kurisu, F., Kasuga, I., Furumai, H., 2016. Molecular characterization of low molecular weight dissolved organic matter in water reclamation processes using orbitrap mass spectrometry. Water Res. 100, 526–536. https://doi.org/10.1016/j.watres.2016.05.047.
Phungsai, P., Kurisu, F., Kasuga, I., Furumai, H., 2018. Changes in dissolved organic matter composition and disinfection byproduct precursors in advanced drinking water treatment processes. Environ. Sci. Technol. 52 (6), 3392–3401. https://doi.org/10.1021/acs.est.7b04765.
Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. J. Cheminf. 8, 3. https://doi.org/10.1186/s13321-016-0115-9.
Ruttkies, C., Neumann, S., Posch, S., 2019a. Improving MetFrag with statistical learning of fragment annotations. BMC Bioinf. 20, 376. https://doi.org/10.1186/s12859-019-2954-7.
Ruttkies, C., Neumann, S., Posch, S., 2019b. Improving MetFrag with statistical learning of fragment annotations. BMC Bioinf. 20.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48 (4), 2097–2098. https://doi.org/10.1021/es5002105.
Shao, L., Deng, Y., Qiu, J., Zhang, H., Liu, W., Baziene, K., Lu, F., He, P., 2021. DOM chemodiversity pierced performance of each tandem unit along a full-scale "MBR+NF" process for mature landfill leachate treatment. Water Res. 195, 117000. https://doi.org/10.1016/j.watres.2021.117000.
Shi, Q., Wu, J., 2021. Review on sulfur compounds in petroleum and its products: State-of-the-art and perspectives. Energy & Fuels 35 (18), 14445–14461. https://doi.org/10.1021/acs.energyfuels.1c02229.
Srivani, P., Sastry, G.N., 2009. Potential choline kinase inhibitors: A molecular modeling study of bis-quinolinium compounds. J. Mol. Graph. Model. 27 (6), 676–688. https://doi.org/10.1016/j.jmgm.2008.10.010.
Vanini, G., Barra, T.A., Souza, L.M., Madeira, N.C.L., Gomes, A.O., Romao, W., Azevedo, D.A., 2020. Characterization of nonvolatile polar compounds from Brazilian oils by electrospray ionization with FT-ICR MS and Orbitrap-MS. Fuel 282, 118790. https://doi.org/10.1016/j.fuel.2020.118790.
Wang, X., Chen, H., Lei, K., Sun, Z., 2015. UVA illumination-induced optical coupling between tryptophan and natural dissolved organic matter. Environ. Sci. Pollut. Res. 22 (21), 16969–16977. https://doi.org/10.1007/s11356-015-4906-8.
Wen, L., Yang, F., Li, X., Liu, S., Lin, Y., Hu, E., Gao, L., Li, M., 2023. Composition of dissolved organic matter (DOM) in wastewater treatment plants influent affects the efficiency of carbon and nitrogen removal. Sci. Total Environ. 857, 159541. https://doi.org/10.1016/j.scitotenv.2022.159541.
Wünsch, U.J., Murphy, K.R., Stedmon, C.A., 2017. The one-sample PARAFAC approach reveals molecular size distributions of fluorescent components in dissolved organic matter. Environ. Sci. Technol. 51 (20), 11900–11908. https://doi.org/10.1021/acs.est.7b03260.
Xie, Y., Fan, H., Che, M., Liu, Y., Liu, C., Hu, X., Teng, B., 2024. Hydrophobicity and pore structure: unraveling the critical factors of alcohol and acid adsorption in zeolites. Molecules 29 (22), 5251. https://doi.org/10.3390/molecules29225251.

Xu, Y., Zhang, Y., Qiu, L., Zhang, M., Yang, J., Ji, R., Vione, D., Chen, Z., Gu, C., 2024. Photochemical behavior of dissolved organic matter in environmental surface waters: A review. EEH 3 (4), 529–542. https://doi.org/10.1016/j.eehl.2024.06.002.

Zhang, B., Shan, C., Wang, S., Fang, Z., Pan, B., 2021. Unveiling the transformation of dissolved organic matter during ozonation of municipal secondary effluent based on FT-ICR-MS and spectral analysis. Water Res. 188, 116484. https://doi.org/10.1016/j.watres.2020.116484.

Zhao, P., Du, Z., Fu, Q., Ai, J., Hu, A., Wang, D., Zhang, W., 2023. Molecular composition and chemodiversity of dissolved organic matter in wastewater sludge via Fourier transform ion cyclotron resonance mass spectrometry: Effects of extraction methods and electrospray ionization modes. Water Res. 232, 119687. https://doi.org/10.1016/j.watres.2023.119687.

Zheng, L., Zhang, T., Gong, R., Xu, S., Zhang, X., Xiao, F., 2025. Identifying the impacts of photochemical and biological processes on wastewater effluent organic matter in receiving water using directed paired mass distance. J. Environ. Manag. 13 (4), 117411. https://doi.org/10.1016/j.jece.2025.117411.