Marine Pollution Bulletin 226 (2026) 119333

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbul

Automated non-target screening workflow for source tracking of organic contaminants in industrialized coastal sediments☆

Mangong Shin [a], Sori Mok [a], Seongjin Hong [b], Hoon Choi [c], Moonjin Lee [c], Hyo-Bang Moon [a,*]

[a] Department of Marine Science and Convergence Technology, College of Engineering Sciences, Hanyang University, Ansan, 15588, Republic of Korea
[b] Department of Earth, Environmental & Space Sciences, Chungnam National University, Daejeon, 34134, Republic of Korea
[c] Maritime Safety and Environment Research Center, Korea Research Institute of Ships and Ocean Engineering, Daejeon, 34103, Republic of Korea

ARTICLE INFO

Keywords:
Non-target
Workflow
Fingerprint
Time-of-flight
Discriminant analysis

ABSTRACT

This study describes the development of an automated workflow for non-target screening data processing and applies multivariate statistical analysis to characterize chemical fingerprints in sediment from industrialized coastal regions. R-based software facilitates streamlined processing of gas chromatography coupled with time-of-flight mass spectrometer datasets through retention index-based filtering, feature consolidation, and blank subtraction, enhancing data reliability and minimizing manual workloads. The workflow is applied to sediment from industrialized bays, and partial least squares discriminant analysis is used to distinguish site-specific chemical fingerprints. Tentatively identified compounds are linked to specific industrial activities, including petrochemical manufacturing, fossil-fuel combustion, waste incineration, and shipbuilding. Model performance and compound relevance are validated through cross-validation and permutation testing, confirming the robustness of the workflow. This integrated approach provides a scalable and reproducible tool for the tracking of chemical sources and environmental monitoring, highlighting the utility of combining non-target screening workflow with statistical analysis in sediment studies.

1. Introduction

Marine sediment serves as both a dynamic component and long-term sink in aquatic environments, frequently accumulating anthropogenic contaminants discharged directly or indirectly into coastal systems (Lee et al., 2020; Kim et al., 2023; Lee et al., 2023). Since the onset of large-scale commercial production of anthropogenic chemicals, numerous organic compounds have been released into coastal environments (Ma et al., 2024). Although many of these compounds occur at low concentrations (nanograms per liter or gram), they can be persistent and bioaccumulate, and some may pose significant ecological risks through acute or chronic toxicity in aquatic organisms (Tian et al., 2020). Conventional analytical methods typically target a limited number of known compounds, potentially overlooking important chemical indicators that are not pre-selected for analysis. Comprehensive analytical approaches are needed to screen for non-targeted contaminants to obtain a holistic understanding of contamination trends in the environment (Wang et al., 2020; Mazur et al., 2021; Ma et al., 2024).

Gas chromatography (GC) coupled with time-of-flight mass spectrometry (TOF/MS) is a powerful analytical tool for the screening of non-target contaminants, enabling the detection and identification of a wide range of known and unknown organic pollutants in environmental samples (Lee et al., 2019; Mok et al., 2023). With GC-TOF based non-target screening, multiple compounds can be identified and characterized using available and accurate mass spectral libraries, potentially reducing the time and cost of acquiring reference standards for compounds that may not be detected in samples (Moschet et al., 2017). However, the large datasets generated by non-target screening require extensive post-processing to extract reliable features, a task that can take weeks or months, preventing the efficient analysis of complex data (Stratton et al., 2024). In some cases, this limitation can slow the progress of chemical experimentation and/or delay analysis and publication.

Effectively processing large-scale non-target screening datasets requires automated workflows that can reliably extract meaningful features and streamline analysis. However, many of the tool currently available either originate from other research domains, such as the life sciences (Aggio et al., 2011; Nodzenski et al., 2014), or depend heavily

☆ This article is part of a Special issue entitled: 'Source Apportionment' published in Marine Pollution Bulletin.

* Corresponding author.
E-mail address: hbmoon@hanyang.ac.kr (H.-B. Moon).

https://doi.org/10.1016/j.marpolbul.2026.119333
Received 4 August 2025; Received in revised form 14 January 2026; Accepted 25 January 2026
Available online 28 January 2026

on parameters sensitive to experimental conditions, limiting their reproducibility and consistency (Helmus et al., 2021). Automated data processing workflows designed for non-target screening can overcome these obstacles. By minimizing dependence on experimental conditions and integrating robust filtering strategies, such workflows improve the reproducibility of results and enhance the accuracy of feature extraction from diverse datasets.

Environmental forensics has attracted increasing attention as a valuable tool for understanding pollution patterns and tracing the sources of contamination. One of the primary objectives of chemical forensic analysis is the identification of diagnostic chemical features, often referred to as chemical fingerprints, which can reliably indicate specific sources of contamination. These fingerprints serve as essential tools for source apportionment, particularly in complex environments with multiple anthropogenic inputs (Zheng et al., 2013; Liu et al., 2025). A previous study used non-target screening to identify organic chemical indicators for tracing contamination sources in groundwater samples (Ekpe et al., 2024). However, to the best of our knowledge, applications involving the use of non-target screening to characterize chemical fingerprints in marine sediments with complicated contamination histories remain limited (Dávila-Santiago et al., 2022). The objectives of this study were to develop an automated workflow for efficiently processing non-target screening data, and to apply this workflow to the identification of contamination sources using chemical fingerprints in sediment from industrialized coastal regions and multivariate statistical analysis. The proposed workflow and accompanying statistical analysis are expected to facilitate source identification of organic contaminants in marine sediment from industrialized coastal environments.

2. Materials and methods

2.1. Sample collection

Sediment samples were collected at 31 sites in Okpo Bay (OP1–OP31) and 26 sites in Onsan Bay (OS1–OS26), both on the southeastern coast of Korea, in May 2023 (Fig. 1). Samples were obtained using a Van Veen grab sampler and stored immediately in solvent-washed aluminum containers. The collected samples were transported to the laboratory in a frozen state at −20 °C, freeze-dried, and homogenized prior to analysis.

Okpo Bay, which hosts large commercial harbors, industrial complexes, and shipyards, has previously been reported to contain high concentrations of polybrominated diphenyl ethers (PBDEs), indicating the environmental pressures associated with industrial activities in the region (Ramu et al., 2010). Onsan Bay, which is home to a variety of industrial complexes and chemical plants, has similarly been affected by severe contamination with persistent organic pollutants (POPs), including PBDEs and polychlorinated naphthalenes (Moon et al., 2007; Lee et al., 2020; Lee et al., 2023), as well as emerging contaminants such as siloxanes and alternative plasticizers (Lee et al., 2018; Kim et al., 2021). These observations prompted the introduction of continuous monitoring of organic pollutants in industrialized coastal areas. As a result, these locations are ideal sampling regions for the identification of chemical fingerprints and source apportionment of environmental contaminants.

2.2. Sample preparation

Sample preparation procedures were similar to those reported previously (Lee et al., 2014; Lee et al., 2020). Briefly, freeze-dried samples of sediment (~10 g) were extracted with accelerated solvent extraction (ASE; Dionex ASE 350, Sunnyvale, CA, USA) using a mixture of dichloromethane (DCM) and hexane (3:1; ultra-trace-residue analysis grade, J.T. Baker, Phillipsburg, NJ, USA). Extracts were concentrated to ~5 mL using a TurboVap LV (Biotage, Uppsala, Sweden) and activated copper was added to remove elemental sulfur. The extract was cleaned using a multi-layer silica gel column with 15% DCM in hexane, which is similar to a method to treat neutral POPs, such as polychlorinated biphenyls (PCBs) and PBDEs. Eluants were concentrated and dissolved in nonane (pesticide analysis grade; Kanto Chemical, Tokyo, Japan) to 100 μL for instrumental analysis. Procedural blanks were prepared using anhydrous sodium sulfate (Kanto Chemical Co., Inc., Tokyo, Japan) and processed identically to the sediment samples.

2.3. Instrumental analysis and non-target analysis

Instrumental analysis for non-target analysis was conducted following previously reported procedures with minor modifications (Lee et al., 2019; Mok et al., 2023; Mok et al., 2024). Extracts were analyzed using a GC-TOF/MS apparatus (Agilent 8890/7250, Agilent Technologies, Santa Clara, CA, USA) equipped with a DB-5MS UI column (30 m length, 0.25 mm inner diameter, 0.25 μm film thickness; J&W Scientific, Palo Alto, CA, USA) in positive electron-ionization mode at 70 eV. An injection volume of 0.5 μL of each final extract was introduced into the GC-TOF/MS using split injection mode (split ratio: 5:1). The oven temperature was programmed to hold at 80 °C for 1 min, ramp to 200 °C at

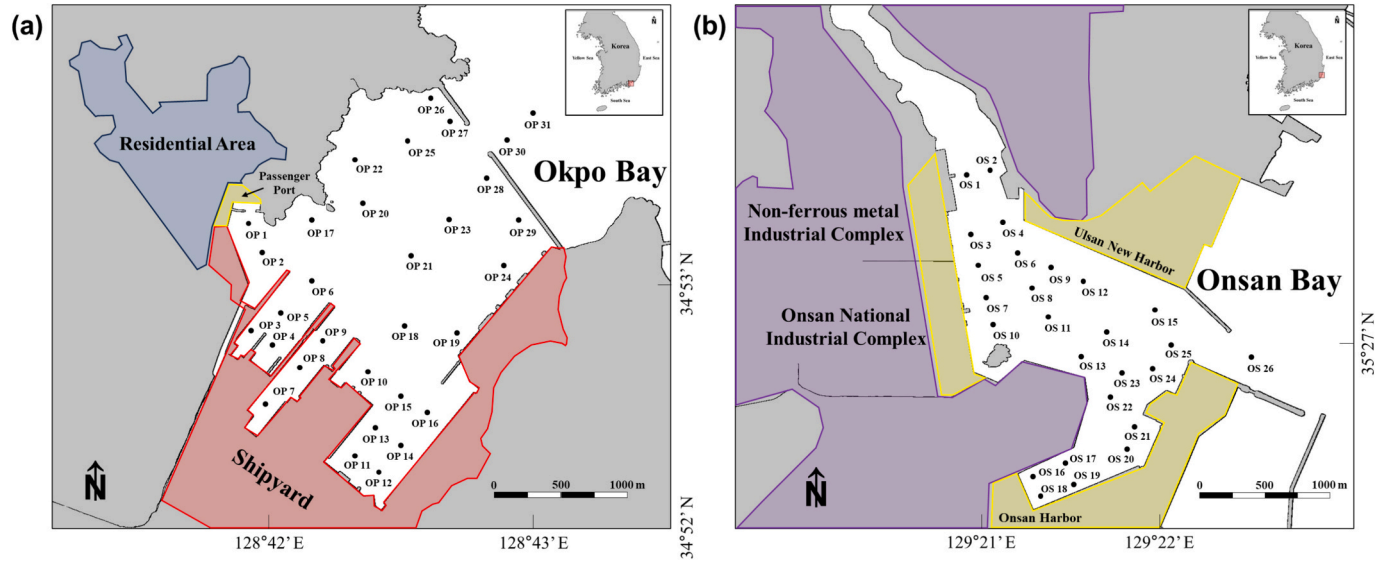

Fig. 1. Sampling locations of sediment collected from (a) Okpo Bay and (b) Onsan Bay of Korea.

10 °C/min, then to 300 °C at 5 °C/min, and hold for 5 min. Helium was used as the carrier gas at a constant flow of 1 mL/min. The temperatures of the transfer line and ion source were 300 °C and 230 °C, respectively. The scan ranged from 50 to 800 mass-to-charge (m/z) at 3 spectra/s. The resolution was approximately 26,000 at m/z 131 and 34,000 at m/z 502 under a 5 GHz high-resolution mode.

Data processing was carried out using Agilent Unknown Analysis software (version 12.0). A deconvolution Sure Mass algorithm was used to extract non-target features from the raw data with retention-time window factors of 50, 100, 200, 300. The minimum to maximum number of ion peaks was 3 to 10. The mass spectra of the features with a peak shape quality >60 were compared to spectra from the National Institute of Standards and Technology (NIST) 17 library to identify unknown compounds. Compounds with a library-matching factor > 60 were retained. Features detected in procedural blanks with solvent were subtracted to eliminate background interference.

2.4. Automated workflow for non-target screening data processing

An automated non-target screening data processing application was developed using R (version 4.2.2) running in RStudio (version 2022.12.0.353) (R core team, 2022; Posit team, 2022). A graphical user interface (GUI) was implemented with the Shiny package (version 1.9.1) (Chang et al., 2024), while dplyr (version 1.1.4) (Wickham et al., 2023) was used for data manipulation and DT (version 0.33) (Xie et al., 2024) for interactive table rendering. The software was optimized to handle raw .csv files exported from Agilent Unknowns Analysis Software, where each file represents a single sample and includes variables such as compound name, Chemical Abstracts Service registry number (CASRN), molecular formula, component area, match factor, library retention index (RI) and delta RI (i.e., the absolute difference between the component RI and the corresponding library RI). All files were saved using sample-specific filenames to enable batch processing. The automated non-target screening workflow consisted of four core modules: (1) RI-based filtering, (2) duplicate feature consolidation, (3) blank integration, and (4) blank subtraction from the samples. RIs for compound identification were calculated based on an injected n-alkane series prior to sample analysis. The n-alkane series served as reference compounds for RI calculation and enabled filtering and more accurate identification of candidate compounds. The RI system, originally proposed by Kováts (1958), is widely used in GC-based analyses as a robust, experiment-independent metric to support compound identification (Ulrich et al., 2013; Babushok, 2015; Ieda et al., 2019). As the first filtering step, only tentative compounds with delta RIs within a user-defined percentage (default: 10%) of the corresponding library RI were retained for further analysis. According to a previous study (Moschet et al., 2018), RI deviations within ±10% are acceptable. A comparison performed between the calculated RI of each non-target feature and the estimated RI (on the n-alkane scale) of candidate compounds in the NIST 17 library. To address the absence of RI values for several candidate compounds in the NIST 17 library, complementary strategies were implemented to estimate and supplement the missing values. One such strategy involved creating a custom "RI_Database.csv" file, allowing users to manually define RI values for unregistered compounds. In this study, predicted RI values for compounds without library RI entries were obtained using GCMS-ID, an open-access RI-prediction platform trained on experimental data using a graph neural network (Wakoli et al., 2024). Specifically, chemical structures (e.g., SMILES) of compounds were submitted to GCMS-ID to obtain predicted RI values, which were then incorporated into the user-defined RI_Database and used to complete the filtering process. When neither library RI nor supplemented RI values (user-defined or structure-based predicted) were available, delta RI could not be evaluated. Therefore, those candidates were flagged and excluded from the RI-based filtering step. A separate script in R and an example format of a user-defined "RI_Database.csv" (provided in Text S1 and Table S1 of the supplementary materials), allowed users to batch-fill missing library RI values across . csv files by matching CASRN and molecular formulas against entries in the custom RI database. This step was not part of the automated software, but it provided a preparatory solution to enhance RI-based filtering. Second, features sharing identical CASRNs and molecular formulas within the same sample file were considered duplicated annotations generated during spectral deconvolution. When multiple entries were detected, a representative feature was selected by prioritizing (i) the smallest absolute delta RI (i.e., best RI agreement) and (ii) the highest library match factor (Perez-Sanz et al., 2021). This consolidation step was implemented to reduce redundancy and improve consistency in downstream statistical analysis. However, it may not fully resolve positional isomers or co-eluting isomers that share the same molecular formula and similar EI mass spectra, and such candidates may be retained as a single representative entry. Third, blank files were merged by averaging the component areas of matching compounds detected across blank replicates. Fourth, procedural blank signals were used to remove background contaminants. Features in the sample files were retained only if their intensity was at least twice the corresponding average in the blank samples – a threshold that can be adjusted by the user – and optionally corrected by subtracting the blank intensity (Ng et al., 2020; Pereira et al., 2021). Following the confidence-level criteria outlined by Schymanski et al. (2014), these candidates were considered tentatively identified compounds at a confidence-level of 2 (Moschet et al., 2018).

The interface was designed to ensure accessibility and ease of use for non-experts (Fig. S1). Upon launching the application, users were presented with a web-based dashboard featuring a sidebar menu and a central display panel. The sidebar gave users the ability to specify the working directory, set numerical parameters, such as RI tolerance and blank removal thresholds, and initiate each step of the workflow using clearly labeled action buttons (e.g., "Run RI Filtering", "Run Blank Removal"). Processed files were automatically saved to designated subfolders, and log files summarizing filtering outcomes were generated to support transparency and reproducibility. The modular structure of the interface enabled partial execution of the workflow, allowing users to re-run specific steps without repeating the entire process. This GUI-based architecture facilitated streamlined, reproducible, and customizable analysis of non-target screening data, even for users without programming experience. This workflow substantially reduced the manual burdens associated with curating non-target screening data and enhanced consistency in downstream statistical and chemometric analysis. The full source code is available in the supplementary materials (Text S2). The application can be run on standard Windows/macOS/Linux systems with R (version 4.2.2 or later) and RStudio installed, without the need for specialized hardware or proprietary software. Although this study implemented the tool as a local GUI to ensure compatibility with restricted laboratory environments, the application is built on the Shiny framework and could be deployed on a server (e.g., Shiny Server) or cloud-based platforms, enabling multi-user access and remote execution in future implementations.

2.5. Data imputation and multivariate statistical analysis for chemical fingerprinting

To address left-censored values in the dataset, we employed MetImp version 1.2, an open-source R-based web tool (https://metabolomics.cc.hawaii.edu/software/MetImp/) (Wei et al., 2018a; Wei et al., 2018b) to implement GSimp (Gibbs sampler-based imputation). GSimp accounts for the uncertainty of missing values during imputation, reducing bias in downstream analyses and improving the accuracy of cross-sample comparisons. This method has outperformed alternative approaches, particularly in datasets with up to 60% censored values, achieving relatively low normalized root mean square errors (Wang et al., 2023). A comprehensive explanation of the GSimp algorithm is available in Wei et al. (2018a).

For multivariate statistical analysis, MetaboAnalyst 6.0 (https://www.metaboanalyst.ca/) (Xia and Wishart, 2011), another R-based web platform, was used. Partial least squares discriminant analysis (PLS-DA) was used to discriminate among chemical fingerprints, and the Kruskal–Wallis test was used for univariate comparisons. A p-value less than 0.05 was considered statistically significant. The most discriminating compounds were identified based on their variable importance in projection (VIP) scores, which were calculated as follows:

$$VIP_i = \sqrt{\frac{p\sum_{a=1}^{A} w_{ia}^2 SSa}{\sum_{a=1}^{A} SS_a}}$$

where p is the number of variables, w_{ia} is the weight of variables i on component a, and SS_a represents the sum of squares explained by component a. PLS-DA was employed for supervised discrimination between predefined groups and identification of source-specific chemical fingerprints from high-dimensional non-target screening data. PLS-DA was selected because the primary objective of this study was to discriminate predefined site groups and identify discriminating marker compounds, whereas principal component analysis is primarily used as an unsupervised exploratory tool and does not explicitly maximize separation between predefined groups (Piotrowski and Place, 2019; Dress et al., 2025). Although orthogonal PLS can improve interpretability by separating predictive and orthogonal variation (Forsgren et al., 2025), supervised discriminant models inherently carry a risk of overfitting in high-dimensional and sparse datasets. Therefore, to maintain a conservative and robust modeling strategy, we used PLS-DA, and model robustness was evaluated using five-fold cross-validation and permutation testing (Szymańska et al., 2012). Nevertheless, since PLS-DA is a supervised method, a risk of overfitting may still remain despite validation procedures; therefore, the results should be interpreted cautiously, particularly in the absence of an external validation dataset.

3. Results and discussion

3.1. Validation of the developed automated non-target screening workflow

To validate the developed automated non-target screening workflow, a pooled sample prepared as a synthetic mixture of multiple known organic pollutants, including phthalates, alternative plasticizers, organophosphate esters, and organophosphite antioxidants, was analyzed. The sample was prepared to reflect a wide range of physicochemical properties and retention times, simulating complex environmental mixtures. The workflow was evaluated based on its ability to correctly identify target compounds included in the mixture. It successfully detected 89% of the expected compounds, with identification results closely matching the known standards included in the pooled sample (Table S2). These findings suggest the developed workflow can accurately and reliably detect known substances in complex mixtures.

Following this validation, the developed workflow was further evaluated by applying it to raw GC-TOF/MS data from a recently published study (Mok et al., 2024) to compare the results with those produced by manual identification under similar criteria. The developed workflow successfully identified most of the compounds manually identified in the reference study, missing only one substance (Table S3). Several additional compounds not included in the original publication were newly identified using the developed workflow (Table S4). These findings imply an improved detection capability of the developed workflow and a potential to uncover previously overlooked compounds. While the results highlight enhanced sensitivity and reliability in non-target screening, further validation with larger datasets is recommended to confirm its applicability across diverse analytical conditions.

3.2. Identification of organic contaminants and their distribution characteristics

The workflow for non-target screening developed in this study is presented in Fig. 2. Deconvolution of the GC/TOF-MS chromatograms produced approximately 1070–3300 non-targeted features per sediment sample. Among them, approximately 330–800 features per sample were matched to candidate compounds in the NIST library, using criteria of a peak shape quality >60 and a match factor > 60. Applying further filtering based on the developed workflow (RI filtering with a ± 10% deviation threshold) narrowed the results to 82–259 and 121–224 features in Okpo Bay and Onsan Bay, respectively. Following removal of duplicate entries based on CASRN and molecular formulas, 58–184 features remained in Okpo Bay and 85–158 in Onsan Bay. In the final step, background contaminants were excluded by subtracting compounds detected in procedural blanks, resulting in 15–135 features in Okpo Bay and 31–124 in Onsan Bay. Overall, 1184 unique organic compounds were identified using the combined approach of NIST library-matching and the automated non-target screening workflow developed in our study.

These compounds were categorized into 16 chemical superclasses using ClassyFire-based taxonomy, revealing distinct distribution patterns of organic compounds in sediments depending on industrial characteristics (Feunang et al., 2016). A full list of the identified compounds, including their classification and detection profiles, is provided in Table S5. The identified compound groups consist of benzenoids (n = 415); hydrocarbons (n = 131); hydrocarbon derivatives (n = 4); lipids and lipid-like molecules (n = 102); organic acids and derivatives (n = 79); organic nitrogen compounds (n = 23); organic oxygen compounds (n = 106); organohalogen compounds (n = 20); organoheterocyclic compounds (n = 176); organometallic compounds (n = 37); organosulfur compounds (n = 8); organic 1,3-dipolar compounds (n = 4); phenylpropanoids and polyketides (n = 73); lignans, neolignans, and related compounds (n = 2); alkaloids and derivatives (n = 2); and others (n = 2). Among them, 742 compounds were identified in sediment from Okpo Bay at a detection frequency of 3.2%–90% and 759 compounds were identified in sediment from Onsan Bay at a detection frequency of 3.8%–100%.

The distribution characteristics of each chemical superclass were assessed by calculating the relative contribution of the average peak area of each group to the total average peak area across all identified superclasses, generating site-specific patterns (Fig. 3). In Okpo Bay, the predominant chemical groups were hydrocarbons (36%), lipids and lipid-like molecules (26%), and benzenoids (15%), collectively accounting for more than 77% of the total detected intensity. The high abundance of hydrocarbons and lipid-like substances suggests strong petrochemical activities and organic matter inputs, potentially originating from ship-building and/or mechanical workshops (Luo et al., 2023; Soman et al., 2024). In addition, organic oxygen compounds (6.1%), organoheterocyclic compounds (4.0%), organic acids and derivatives (3.2%), and phenylpropanoids and polyketides (3.1%) made moderate contributions, suggesting inputs from a combination of anthropogenic activities (e.g., industrial discharges and wastewater effluents) and/or natural sources, such as plant-derived metabolites or microbial degradation products. Organohalogens, which are commonly linked to industrial chemicals and flame retardants, accounted for 4.3% of the total intensity, suggesting possible inputs from polymer manufacturing and synthetic chemical applications (Lee et al., 2020). Onsan Bay exhibited a markedly different chemical profile, with benzenoids overwhelmingly dominating (72%), followed by hydrocarbons (7.3%). The predominance of benzenoids indicates substantial input of aromatic hydrocarbons, consistent with the presence of large-scale petrochemical and chemical manufacturing facilities in the area (Lee et al., 2020; Kim et al., 2023; Lee et al., 2023). This may reflect ongoing emissions of aromatic compounds or their derivatives, possibly related to solvent use, high-temperature combustion, and/or incomplete

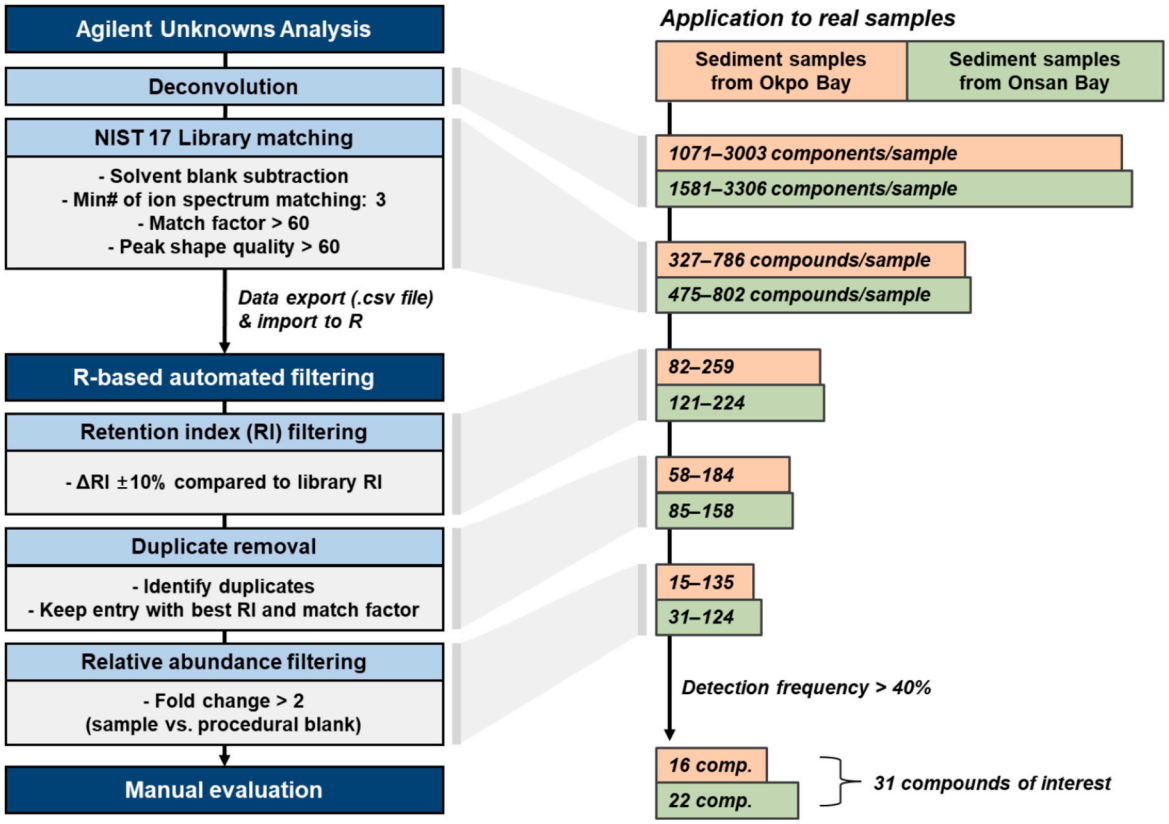

Fig. 2. Automated workflow for non-target screening data and compound filtering based on GC/TOF-MS data.

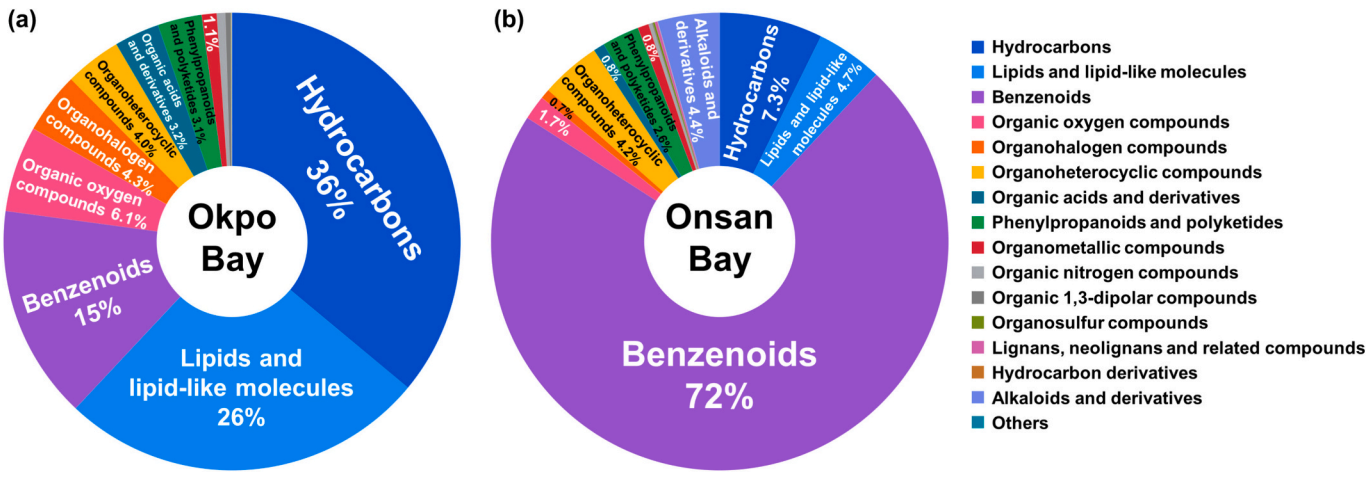

Fig. 3. Group-wised distribution of chemical superclasses identified in sediment from (a) Okpo Bay and (b) Onsan Bay of Korea.

combustion. In addition, lipids and lipid-like molecules (4.7%), alkaloids and derivatives (4.4%), and organoheterocyclic compounds (4.2%) were also present at moderate levels, indicating diverse anthropogenic and/or natural sources. Chemical classification was performed using ClassyFire, a widely used tool for structure-based chemical taxonomy. While effective for high-level categorization, this approach could be limited in that it considers only the molecular structure, and potentially omits contextual information, such as specific usage patterns and environmental behavior. Moreover, certain compounds may belong to multiple classes or functional groups, and more detailed interpretation commonly requires reviewing each compound individually (Aurich et al., 2023).

3.3. Multivariate statistical analysis for source tracking

Mass spectrometry based environmental and metabolomics datasets commonly contain numerous zeros or "missing values" due to low analyte concentrations in the sample matrix. If these left-censored data are not properly managed prior to statistical analysis, they can result in bias that distorts variable correlations and complicates the identification of marker compounds (Hrydziuszko and Viant, 2012; Yang et al., 2015). To address this issue, a series of preprocessing steps was applied, including both filtering and robust imputation. Given the generally low detection frequency of compounds in sediment samples, a 40% group-wise detection-frequency cut-off was used. Compounds detected in at least 40% of samples in any one group were retained. As a result of this

preprocessing, the number of variables was reduced from 1184 to 31, which were then used for multivariate statistical analysis. The list of retained compounds, their abbreviations, molecular formulas, retention indices, CASRN, and detection frequencies in sediments from Okpo and Onsan Bays are provided in Table 1.

The PLS-DA model was used to identify representative indicators for each source (Antignac et al., 2016; Ekpe et al., 2024). To address differences in scale among variables, a log transformation followed by auto-scaling was applied. This preprocessing ensured that high-intensity compounds contributed meaningfully to explanations of variance without disproportionately dominating the results. To reduce the risk of overfitting, five-fold cross-validation was applied. In this procedure, the dataset was randomly divided into five approximately equal subsets. For each iteration, one subset was reserved for testing the model, while the remaining four were used for training. The PLS-DA model demonstrated strong explanatory and predictive performance, with a coefficient of determination of 0.826, a predictive ability of 0.773, and a classification accuracy of 98.2% (Fig. S2), indicating good model fit and minimal risk of overfitting. A two-dimensional score plot of the PLS-DA model revealed clear separation of sediment samples based on their chemical source profiles (Fig. S3), with the first two components explaining 46% of the total variance. VIP scores were used to identify the compounds most responsible for group separation. Compounds with VIP scores $\geq$0.7 were considered discriminative, yielding 22 representative substances (Fig. 4). Each of these compounds showed significantly higher intensities in least one group, as validated by Kruskal–Wallis tests ($p < 0.05$) (Fig. S4). To further assess the statistical robustness of the model, permutation testing was conducted using an approach described by Szymańska et al. (2012). The model's classification performance was compared against 2000 randomly permuted datasets. The original model outperformed all permuted models ($p < 0.0005$), confirming the reliability of the 22 selected chemical fingerprints (Fig. S5).

3.4. Suggestion of chemical indicators predicted through multivariate statistical analysis

The chemical indicators predicted by the PLS-DA and their known applications or uses are suggested in Table 2. Among the compounds identified in sediments from Okpo Bay, several alkylated indane derivatives were specifically enriched, including 1 h-Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl- (TPM-Ind), 1 h-Indene, 2,3-dihydro-1,1,2,3,3-pentamethyl- (PM-Ind), and 1,1′-spirobiindene, 3,3,3′,3’(2H,2H’)-tetramethyl- (TMSpiInd). These compounds belong to the indanes class and have been reported in various industrial and consumer applications. TPM-Ind has been previously detected in fragrance formulations, liquid resins, and recycled high-density polyethylene, suggesting its potential origin from plastic reprocessing, resin manufacturing, and consumer product usage (Na et al., 2022; Woods et al., 2007; Zhang et al., 2025). PM-Ind, while also reported in natural matrices such as plant leaves (Al-Otibi et al., 2022), shares structural similarities with known fragrance ingredients (US EPA CompTox Chemicals Dashboard; https://comptox.epa.gov/dashboard; retrieved 06/2025) (Williams et al., 2017),

Table 1
List of retained compounds for statistical analysis with abbreviations, molecular formulas, retention time, CAS registry number, and detection frequencies in sediment from Okpo and Onsan Bays of Korea.

	Retention time	Compound name	Abbreviations	Formula	CAS Registry Number	Fold change[a]	Detection frequency of compounds (%)	
							Okpo Bay ($n = 31$)	Onsan Bay ($n = 26$)
1	3.73	Heptane, 2,2,4,6,6-pentamethyl-	PM-Hep	$C_{12}H_{26}$	13,475–82-6	2.4	32	42
2	4.37	3-Heptene, 2,2,3,5,6-pentamethyl-	PM-Hept	$C_{12}H_{24}$	116,164–06-8	3.8	55	35
3	5.48	Benzene, 1,2,4,5-tetramethyl-	TMBz	$C_{10}H_{14}$	95–93-2	4.9	26	46
4	6.42	Naphthalene	Nap	$C_{10}H_{8}$	91–20-3	2.3	39	69
5	6.52	Nonane, 3-methyl-5-propyl-	3M5P-Non	$C_{13}H_{28}$	31,081–18-2	2.0	48	50
6	7.42	Naphthalene, decahydro-2,3-dimethyl-	DH-DM-Nap	$C_{12}H_{22}$	1008-80-6	7.4	26	65
7	8.22	1H-Indene, 2,3-dihydro-1,1,2,3,3-pentamethyl-	PM-Ind	$C_{14}H_{20}$	1203-17-4	13	55	12
8	9.42	1,4-Dimethyl-2-cyclopentylbenzene	DM-CPBz	$C_{13}H_{18}$	62,379–92-4	18	61	73
9	12.81	Methane, di-p-tolyl-	DiPTol	$C_{15}H_{16}$	4957-14-6	–	35	46
10	13.30	1H-Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl-	TPM-Ind	$C_{18}H_{20}$	3910-35-8	–	45	ND[c]
11	13.50	Benzene, (1-butyloctyl)-	BuOBz	$C_{18}H_{30}$	2719-63-3	–	10	42
12	15.59	*o*-Terphenyl	o-TP	$C_{18}H_{14}$	84–15-1	–	3.2	92
13	16.27	Benzene, 1,1′-(1,1,2,2-tetramethyl-1,2-ethanediyl)bis[4-methyl-	TMEBz	$C_{20}H_{26}$	734–17-8	–	19	42
14	16.42	1,1’-Spirobiindene, 3,3,3′,3’(2H,2H’)-tetramethyl-	TMSpiInd	$C_{21}H_{24}$	58,343–29-6	–	71	ND
15	16.91	Eicosane	C20	$C_{20}H_{42}$	112–95-8	–	68	12
16	17.25	p-Dicyclohexylbenzene	p-DCHBz	$C_{18}H_{26}$	1087-02-1	–	ND	46
17	17.72	Fluoranthene	Flu	$C_{16}H_{10}$	206–44-0	–	45	7.7
18	18.32	Pyrene	Pyr	$C_{16}H_{10}$	129–00-0	–	52	65
19	18.80	Cyclohexane, [1,1′-biphenyl]-4-yl-	BP-Cyclo	$C_{18}H_{20}$	NR[b]	–	ND	100
20	19.17	*m*-Terphenyl	m-TP	$C_{18}H_{14}$	92–06-8	–	3.2	77
21	19.76	*p*-Terphenyl	p-TP	$C_{18}H_{14}$	92–94-4	–	13	50
22	23.02	Triphenylene	Triphe	$C_{18}H_{12}$	217–59-4	–	65	54
23	26.99	Phenanthrene	Phe	$C_{14}H_{10}$	85–01-8	–	48	35
24	27.23	Benzo[*b*]fluoranthene	BbF	$C_{20}H_{12}$	205–99-2	–	58	7.7
25	30.07	1,1′:3′,1”-Terphenyl, 5′-phenyl-	Ph-TP	$C_{24}H_{18}$	612–71-5	–	13	96
26	30.29	17αH-Trisnorhopane	TrisNorHop	$C_{27}H_{46}$	53,584–59-1	–	35	46
27	31.84	*m,p*-Quaterphenyl	mp-QP	$C_{24}H_{18}$	1166-19-4	–	ND	50
28	32.20	28-Nor-17α(H)-hopane	NorHop-28	$C_{29}H_{50}$	53,584–60-4	–	90	85
29	32.26	19-Norcholesta-1,3,5(10)-trien-6-one	NorChol	$C_{26}H_{38}O$	19,454–79-6	–	68	7.7
30	33.25	17α(H),21β(H)-hopane	Hop-17,21	$C_{30}H_{52}$	13,849–96-2	–	81	92
31	36.50	Benzene, 1,1′-[1,2-ethanediylbis(oxy)]bis[2,4,6-tribromo-	BTBPE	$C_{14}H_{8}Br_{6}O_{2}$	37,853–59-1	–	45	65

[a] Values of fold change (blank vs. sample) calculated by dividing average peak area, dash (–) indicates compounds not detected in the blank.
[b] NR: No CAS registry number was reported.
[c] ND: Not detected.

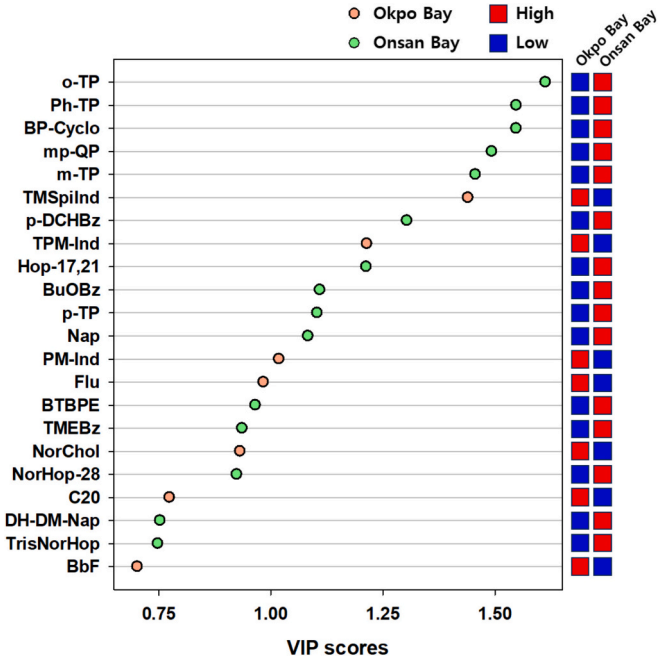

Fig. 4. Cleveland dot plot of variable importance in projection (VIP) scores from the partial least squares discriminant analysis (PLS-DA) model, highlighting key substances (VIP $\geq$ 0.7) associated with pollution source differentiation. Higher VIP scores indicate greater influence on the model. The heatmap on the right shows the relative abundance of each compound across sediment samples from Okpo Bay and Onsan Bay of Korea. Abbreviations: o-TP, *o*-terphenyl; pH-TP, 1,1′:3′,1″-terphenyl, 5′-phenyl-; BP-Cyclo, cyclohexane, [1,1′-biphenyl]-4-yl-; mp-QP, *m,p*-quaterphenyl; m-TP, *m*-terphenyl; TMSpiInd, 1,1′-spirobiindene, 3,3,3′,3′(2H,2H')-tetramethyl-; p-DCHBz, p-dicyclohexylbenzene; TPM-Ind, 1H-indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl-; Hop-17,21, 17α(H),21β(H)-hopane; BuOBz, benzene, (1-butyloctyl)-; p-TP; *p*-terphenyl; Nap, naphthalene; PM-Ind, 1H-indene, 2,3-dihydro-1,1,2,3,3-pentamethyl-; Flu, fluoranthene; BTBPE, benzene, 1,1′-[1,2-ethanediylbis(oxy)]bis [2,4,6-tribromo-; TMEBz, benzene, 1,1′-(1,1,2,2-tetramethyl-1,2-ethanediyl)bis [4-methyl-; NorChol, 19-norcholesta-1,3,5(10)-trien-6-one; NorHop-28, 28-nor-17α(H)-hopane; C20, eicosane; DH-DM-Nap, naphthalene, decahydro-2,3-dimethyl-; TrisNorHop, 17αH-trisnorhopane; BbF, benzo[*b*]fluoranthene.

indicating mixed natural (biogenic) and anthropogenic origins. TMSpiInd, in contrast, has been identified as a constituent of industrial catalyst, pointing to a direct association with manufacturing activities (Lin et al., 2018). In particular, indane-based compounds have been widely reported in crude oil (Berthod et al., 1998), supporting the possibility of petrogenic contamination in sediments from Okpo Bay. This was also evidenced by their high detection frequencies (TPM-Ind: 45%, PM-Ind: 55%, and TMSpiInd: 71%) in sediment from Okpo Bay, and much lower or non-detectable levels in sediment from Onsan Bay. These findings are likely due to the presence of large-scale ship-building and repair facilities in Okpo that may contribute oil-derived emissions from lubricants, fuels, and solvents (Shim et al., 2002; Yokoyama and Choi, 2010). A petrogenic influence on sediment from Okpo Bay was also supported by the detection of eicosane (C20), a straight-chain saturated hydrocarbon commonly used as a marker of fresh oil inputs, particularly diesel and gasoline (Reddy et al., 2007). The co-occurrence of eicosane and indane derivatives strengthens the evidence for petroleum-related contamination. In addition, polycyclic aromatic hydrocarbons (PAHs), such as fluoranthene and benzo[*b*]fluoranthene, were also abundant in the region. PAHs are well-known chemical markers of pyrogenic input that are formed primarily during the incomplete combustion of fossil fuels, and are typically associated with welding, high-temperature operations, and fossil-fuel use (Moon et al., 2011). Last, the detection of 19-norcholesta-1,3,5(10)-trien-6-one, a steroidal compound known as

Table 2
Summary of predicted chemical indicators in this study and their applications/uses. Table 1 supplies the full names of the compounds.

Region	Chemical indicators	Class	Applications/uses
Okpo Bay	TPM-Ind	Alkylated indanes	Fragrance, resin, recycled plastic
	PM-Ind	Alkylated indanes	Natural product (leaves), possible fragrance
	TMSpiInd	Alkylated indanes	Industrial catalyst
	C20 (Eicosane)	n-alkanes	Marker for fresh oil input (diesel/gasoline)
	Flu, BbF	PAHs	Combustion byproduct (fossil fuels)
	NorChol	Steroid derivatives	Suspected to be associated with municipal or hospital effluents
Onsan Bay	Nap	PAHs	Diesel/gasoline combustion product
	Hop-17,21, NorHop-28, TrisNorHop	Hopanes	Diesel/gasoline combustion product
	DH-DM-Nap	PAH derivatives	Coal-based jet fuel
	Ph-TP, o-TP, m-TP, p-TP, mp-QP	Polyphenyls	Waste and e-waste incineration byproducts
	BP-Cyclo	Aromatic hydrocarbon	PET solvent
	p-DCHBz	Aromatic hydrocarbon	Industrial catalyst/additive
	BTBPE	NBFR	Brominated flame retardant
	BuOBz, TMEBz	Aromatic hydrocarbons	Natural products; possible fragrance use

NorChol, suggests an additional potential source of contamination. While NorChol can occur naturally, steroid derivatives as a chemical class are frequently reported in municipal and hospital effluents, indicating a possible pharmaceutical or domestic origin (Ting and Praveena, 2017). Given the residential areas around Okpo Bay, it is plausible that domestic wastewater discharge is contributing to the observed contamination profile. These findings collectively suggest the combined influence of industrial activities and municipal sources in shaping organic contaminant profiles in sediments from Okpo Bay.

Among the compounds identified in sediment from Onsan Bay, a group of hydrocarbons, including naphthalene (Nap) and hopane-type compounds such as 17α(H),21β(H)-hopane (Hop-17,21), 28-nor-17α(H)-hopane (NorHop-28), and 17α(H)-trisnorhopane (TrisNorHop), was indicative of petrogenic input associated with fossil-fuel combustion (Schauer et al., 1999; Schauer et al., 2002). The concurrent detection of these compounds in Onsan Bay sediment suggests contamination from fossil-fuel related sources, likely industrial operations, vehicular emissions, or marine traffic in the surrounding area. In addition, naphthalene, decahydro-2,3-dimethyl- (DH-DM-Nap), a saturated naphthalene derivative, was also detected. This compound has been reported in coal-based jet fuel, suggesting potential inputs from coal-derived liquid fuels used in industrial applications (Smith and Bruno, 2007). A group of polyphenyl compounds, 1,1′:3′,1″-terphenyl, 5′-phenyl- (pH-TP), *o*-terphenyl (o-TP), *m*-terphenyl (m-TP), *p*-terphenyl (p-TP), and *m,p*-quaterphenyl (mp-QP), was also found at enriched levels in Onsan Bay sediment. These compounds are byproducts of high-temperature degradation of synthetic materials and have been reported in emissions from municipal waste incineration and plastic combustion (Hoffer et al., 2021). In particular, pH-TP has been identified in the pyrolysis products of electronic waste, implying contributions from electronic waste disposal or recycling (Sahle-Demessie et al., 2021). Combustion processes at industrial complexes in Onsan Bay have been documented as potential sources of a variety of contaminants (Lee et al., 2023). Two additional aromatic compounds, cyclohexane, [1,1′-biphenyl]-4-yl- (BP-Cyclo) and p-dicyclohexylbenzene (p-DCHBz), were also prevalent. BP-Cyclo is a solvent reportedly used in the synthesis of poly(ethylene

terephthalate) (PET) (Tate et al., 1993), while p-DCHBz functions as a catalyst or additive in industrial chemical processes (Bao et al., 2022). The detection of these compounds suggests inputs from polymer manufacturing or other solvent-intensive industrial activities surrounding Onsan Bay. Another predominantly detected compound was benzene, 1,1′-[1,2-ethanediylbis(oxy)]bis[2,4,6-tribromo- (BTBPE), a well-known novel brominated flame retardant. As a replacement for legacy brominated flame retardants (BFRs) such as PBDEs, BTBPE is commonly incorporated into plastics, textiles, and electronic products. Its hydrophobicity and persistence make it prone to sediment accumulation, and previous studies have reported finding it in both water and sediment from Onsan Bay (Lee et al., 2020). Its detection in this study suggests ongoing or historical discharges from local electronics manufacturing or recycling facilities. Last, benzene, (1-butyloctyl)- (BuOBz) and benzene, 1,1′-(1,1,2,2-tetramethyl-1,2-ethanediyl)bis[4-methyl- (TMEBz) were also detected. These compounds have been reported in natural products, such as roasted coffee, suggesting a potential biogenic origin (John et al., 2021; Paiva et al., 2025). However, due to their structural similarity to fragrance materials, they may also be associated with anthropogenic sources, including emissions from consumer products or food-processing activities (CompTox Chemicals Dashboard; retrieved 06/2025). Overall, the suite of compounds identified in sediment from Onsan Bay reflects a complicated mixture of contamination sources, including fossil-fuel combustion, waste incineration, industrial manufacturing, and possible contributions from consumer product usage.

4. Conclusions and study limitations

An automated workflow of non-target screening data processing and a multivariate statistical tool were effectively developed and combined to characterize chemical fingerprints in sediment from industrialized coastal regions. The automated workflow streamlined the processing of large-scale GC-TOF/MS datasets through RI-based filtering, duplicate removal, and blank subtraction, enabling consistent and reproducible detection of contaminants. Distinct chemical fingerprints observed in sediment from Okpo and Onsan Bays reflected a wide range of pollution sources, including petrochemicals, combustion byproducts, and industrial additives.

Despite this study's strengths, several limitations should be considered. First, the identification of compounds is tentative and relies on spectral library-matching; expert validation is required, particularly for compounds with similar structures or isomeric forms. Second, while the workflow improves efficiency, the duplicate consolidation step (based on CASRN and molecular formula) may group multiple congeners and/or isomeric candidates under a single representative entry, particularly when compounds show similar EI mass spectra and retention behavior, potentially affecting interpretation. This could lead to misinterpretation of halogenated contaminants, such as PCBs. Finally, the absence of authentic standards for many compounds limited our ability to confirm identities and quantify concentrations. Addressing these limitations will enhance the utility of the workflow, not only for environmental monitoring in industrialized coastal regions but also for managing emerging pollution challenges in diverse ecological and regulatory contexts.

CRediT authorship contribution statement

Mangong Shin: Writing – original draft, Visualization, Software, Methodology. **Sori Mok:** Software, Methodology, Investigation, Formal analysis, Data curation. **Seongjin Hong:** Validation, Resources, Methodology, Investigation, Formal analysis, Data curation. **Hoon Choi:** Methodology, Investigation, Formal analysis, Data curation. **Moonjin Lee:** Project administration, Funding acquisition, Formal analysis, Data curation. **Hyo-Bang Moon:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This study was supported by “Development of technology for impact assessment and management of HNS discharged from marine industrial facilities (RS-2021-KS211535)”, “Development of source identification and apportionment methods for toxic substances in marine environments (RS-2022-KS221655)”, and “Development of management technology for persistent organic pollutants (POPs) in the marine environment (RS-2024-00417889)” projects of the Korea Institute of Marine Science and Technology Promotion, which is funded by the Ministry of Oceans and Fisheries, Korea. This work was also supported by a National Research Foundation of Korea grant funded by the Korea government (MSIT) (RS-2025-02263830).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2026.119333.

Data availability

Data will be made available on request.

References

Aggio, R., Villas-Bôas, S.G., Ruggiero, K., 2011. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. Bioinformatics 27, 2316–2318.
Al-Otibi, F.O., Alrumaizan, G.I., Alharbi, R.I., 2022. Evaluation of anticandidal activities and phytochemical examination of extracts prepared from Vitex agnus-castus: a possible alternative in treating candidiasis infections. BMC Complementary Med. Ther. 22, 69.
Antignac, J.P., Main, K.M., Virtanen, H.E., Boquien, C.Y., Marchand, P., Venisseau, A., Guiffard, I., Bichon, E., Wohlfahrt-Veje, C., Legrand, A., Boscher, C., Skakkebæk, N. E., Toppari, J., Le Bizec, B., 2016. Country-specific chemical signatures of persistent organic pollutants (POPs) in breast milk of French, Danish and Finnish women. Environ. Pollut. 218, 728–738.
Aurich, D., Diderich, P., Helmus, R., Schymanski, E.L., 2023. Non-target screening of surface water samples to identify exposome-related pollutants: a case study from Luxembourg. Environ. Sci. Eur. 35, 94.
Babushok, V.I., 2015. Chromatographic retention indices in identification of chemical compounds. Trends Anal. Chem. 69, 98–104.
Bao, R., Shen, Z., Ma, C., Li, L., Wang, Y., Sun, H., Yang, W., 2022. Delaminated MWW-type zeolite and its catalytic performance in alkylation of benzene with cyclohexene. Appl. Catal. A 643, 118741.
Berthod, A., Wang, X., Gahm, K.H., Armstrong, D.W., 1998. Quantitative and stereoisomeric determination of light biomarkers in crude oil and coal samples. Geochim. Cosmochim. Acta 62, 1619–1630.
Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2024. Shiny: web application framework for R. R package version 1 (9), 1. Available online: https://CRAN.R-project.org/package=shiny/.
Dávila-Santiago, E., Shi, C., Mahadwar, G., Medeghini, B., Insinga, L., Hutchinson, R., Good, S., Jones, G.D., 2022. Machine learning applications for chemical fingerprinting and environmental source tracking using non-target chemical data. Environ. Sci. Technol. 56, 4080–4090.
Dress, F., Vosough, M., Schmidt, T.C., 2025. Unraveling non-target screening variability for LC-HRMS data: a chemometric comparative analysis of river water samples impacted by treated wastewater. Anal. Bioanal. Chem. 417, 6075–6088.
Ekpe, O.D., Choo, G., Kang, J.-K., Yun, S.-T., Oh, J.-E., 2024. Identification of organic chemical indicators for tracking pollution sources in groundwater by machine learning from GC-HRMS-based suspect and non-target screening data. Water Res. 252, 121130.
Feunang, Y.D., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S., 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J. Chemother. 8, 61.
Forsgren, E., Björkblom, B., Trygg, J., Jonsson, P., 2025. OPLS-based multiclass classification and data-driven interclass relationship discovery. J. Chem. Inf. Model. 65, 1762–1770.

Helmus, R., ter Laak, T.L., van Wezel, A.P., de Voogt, P., Schymanski, E.L., 2021. patRoon: open source software platform for environmental mass spectrometry based non-target screening. J. Chem. 13, 1.
Hoffer, A., Tóth, Á., Jancsek-Turóczi, B., Machon, A., Meiramova, A., Nagy, A., Marmureanu, L., Gelencsér, A., 2021. Potential new tracers and their mass fraction in the emitted PM10 from the burning of household waste in stoves. Atmos. Chem. Phys. 21, 17855–17864.
Hrydziuszko, O., Viant, M.R., 2012. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. Metabolomics 8, 161–174.
Ieda, T., Hashimoto, S., Isobe, T., Kunisue, T., Tanabe, S., 2019. Evaluation of a data-processing method for target and non-target screening using comprehensive two-dimensional gas chromatography coupled with high-resolution time-of-flight mass spectrometry for environmental samples. Talanta 194, 461–468.
John, S.M., Deeseenthum, S., Luang-In, V., Chottanom, P., 2021. Comparative analysis of volatile compounds and antioxidant activity of kefir produced by thai black jasmine rice. J. Sustain. Sci. Manag. 16, 45–62.
Kim, S., Kim, Y., Moon, H.-B., 2021. Contamination and historical trends of legacy and emerging plasticizers in sediment from highly industrialized bays of Korea. Sci. Total Environ. 765, 142751.
Kim, Y., Hong, S., Jun, L., Lee, Y., Kim, M., Yim, U.H., Khim, J.S., Shin, K.-H., 2023. Use of molecular composition and compound-specific isotope analysis for source appointment of PAHs in sediments of a highly industrialized area. Environ. Pollut. 337, 122546.
Kováts, E., 1958. Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: Retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. Helv. Chim. Acta 41, 1915–1932 (in German).
Lee, H.-H., Lee, S., Lee, M., Moon, H.-B., 2023. Spatial and temporal trends in polychlorinated naphthalenes in sediment from Ulsan and Onsan bays of Korea: potential sources and ecotoxicological concerns. Environ. Geochem. Health 45, 6793–6806.
Lee, H.-K., Lee, S., Lim, J.-E., Moon, H.-B., 2020. Legacy and novel flame retardants in water and sediment from highly industrialized bays of Korea: occurrence, source tracking, decadal time trend, and ecological risks. Mar. Pollut. Bull. 160, 111639.
Lee, S., Song, G.-J., Kannan, K., Moon, H.-B., 2014. Occurrence of PBDEs and other alternative brominated flame retardants in sludge from wastewater treatment plants in Korea. Sci. Total Environ. 470, 1422–1429.
Lee, S., Kim, K., Jeon, J., Moon, H.-B., 2019. Optimization of suspect and non-target analytical methods using GC/TOF for prioritization of emerging contaminants in the Arctic environment. Ecotoxicol. Environ. Saf. 181, 11–17.
Lee, S.-Y., Lee, S., Choi, M., Kannan, K., Moon, H.-B., 2018. An optimized method for the analysis of cyclic and linear siloxanes and their distribution in surface and core sediments from industrialized bays in Korea. Environ. Pollut. 236, 111–118.
Lin, X., Sun, W., GU H., 2018. 3,3,3′,3′-tetramethyl-1,1′-spirobiindane-based phosphinooxazoline ligand compound, preparation method and uses of the same. US Patent US20210079030A1.
Liu, Y., Zhang, L., Wang, Z., Wu, R., Liu, X., Zhu, X., Li, H., Bian, Y., Wang, Z., Zhao, Y., 2025. An inverse dispersion method based on fingerprint species for estimating emission intensities of volatile organic compounds in the chemical industry. ACS ES&T Air 2, 1161–1171.
Luo, Y., Tong, G., Song, Q., Tao, P., Jin, M., Gu, N., Zheng, M., Yu, X., Yu, X., 2023. Impacts of shipyard oil leakage on the PAHs and PCBs occurrence in Xiangshan Bay, China. Mar. Environ. Res. 190, 106057.
Ma, Y., Yang, C., Yao, Q., Li, F., Mao, L., Zhou, X., Meng, X., Chen, L., 2024. Nontarget screening analysis of organic compounds in river sediments: a case study in the Taipu River of the Yangtze River Delta region in China. Environ. Sci. Pollut. Res. 31, 24547–24558.
Mazur, D.M., Detenchuk, E.A., Sosnova, A.A., Artaev, V.B., Lebedev, A.T., 2021. GC-HRMS with complementary ionization techniques for target and non-target screening for chemical exposure: expanding the insights of the air pollution markers in Moscow snow. Sci. Total Environ. 761, 144506.
Mok, S., Lee, S., Choi, Y., Jeon, J., Kim, Y.H., Moon, H.-B., 2023. Target and non-target analyses of neutral per-and polyfluoroalkyl substances from fluorochemical industries using GC-MS/MS and GC-TOF: insights on their environmental fate. Environ. Int. 182, 108311.
Mok, S., Lee, S., Lee, N., Kim, S., Choi, K., Park, J., Kho, Y., Moon, H.-B., 2024. Nationwide human biomonitoring strategy in Korea: prioritization of novel contaminants using GC/TOF-MS with suspect and non-target screening. Chemosphere 369, 143814.
Moon, H.-B., Kannan, K., Choi, M., Choi, H.-G., 2007. Polybrominated diphenyl ethers (PBDEs) in marine sediments from industrialized bays of Korea. Mar. Pollut. Bull. 54, 1402–1412.
Moon, H.-B., An, Y.-R., Park, K.J., Choi, S.-G., Moon, D.-Y., Choi, M., Choi, H.-G., 2011. Occurrence and accumulation features of polycyclic aromatic hydrocarbons and synthetic musk compounds in finless porpoises (*Neophocaena phocaenoides*) from Korean coastal waters. Mar. Pollut. Bull. 62, 1963–1968.
Moschet, C., Lew, B.M., Hasenbein, S., Anumol, T., Young, T.M., 2017. LC-and GC-QTOF-MS as complementary tools for a comprehensive micropollutant analysis in aquatic systems. Environ. Sci. Technol. 51, 1553–1561.
Moschet, C., Anumol, T., Lew, B.M., Bennett, D.H., Young, T.M., 2018. Household dust as a repository of chemical accumulation: new insights from a comprehensive high-resolution mass spectrometric study. Environ. Sci. Technol. 52, 2878–2887.
Na, M., O'Brien, D., Gerberick, G.F., Kern, P.S., Lavelle, M., Lee, I., Parakhia, R., Ryan, C., Api, A.M., 2022. Benchmarking performance of SENS-IS assay against weight of evidence skin sensitization potency categories. Regul. Toxicol. Pharmacol. 130, 105128.
Ng, B., Quinete, N., Gardinali, P.R., 2020. Assessing accuracy, precision and selectivity using quality controls for non-targeted analysis. Sci. Total Environ. 713, 136568.
Nodzenski, M., Muehlbauer, M.J., Bain, J.R., Reisetter, A.C., Lowe Jr., W.L., Scholtens, D. M., 2014. Metabomxtr: an R package for mixture-model analysis of non-targeted metabolomics data. Bioinformatics 30, 3287–3288.
Paiva, A.C., Teixeira, C.A., Hantao, L.W., 2025. Enhancing the understanding of aroma formation during coffee roasting using DHS-GC×GC-TOFMS and chemometrics. ACS Omega 10, 26871–26883.
Pereira, K.L., Ward, M.W., Wilkinson, J.L., Sallach, J.B., Bryant, D.J., Dixon, W.J., Hamilton, J.F., Lewis, A.C., 2021. An automated methodology for non-targeted compositional analysis of small molecules in high complexity environmental matrices using coupled ultra performance liquid chromatography orbitrap mass spectrometry. Environ. Sci. Technol. 55, 7365–7375.
Perez-Sanz, F., Ruiz-Hernández, V., Terry, M.I., Arce-Gallego, S., Weiss, J., Navarro, P.J., Egea-Cortines, M., 2021. gcProfileMakeR: An r package for automatic classification of constitutive and non-constitutive metabolites. Metabolites 11, 211.
Piotrowski, P., Place, B., 2019. Comparison of pre-processing and variable selection strategies in group-based GC×GC-TOFMS analysis. Separations 6, 41.
Posit team, 2022. RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. Available online: http://www.posit.co/.
R core team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/.
Ramu, K., Isobe, T., Takahashi, S., Kim, E.-Y., Min, B.-Y., We, S.-U., Tanabe, S., 2010. Spatial distribution of polybrominated diphenyl ethers and hexabromocyclododecanes in sediments from coastal waters of Korea. Chemosphere 79, 713–719.
Reddy, C.M., Nelson, R.K., Sylva, S.P., Xu, L., Peacock, E.A., Raghuraman, B., Mullins, O. C., 2007. Identification and quantification of alkene-based drilling fluids in crude oils by comprehensive two-dimensional gas chromatography with flame ionization detection. J. Chromatogr. A 1148, 100–107.
Sahle-Demessie, E., Mezgebe, B., Dietrich, J., Shan, Y., Harmon, S., Lee, C.C., 2021. Material recovery from electronic waste using pyrolysis: emissions measurements and risk assessment. J. Environ. Chem. Eng. 9, 104943.
Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, B.R.T., 1999. Measurement of emissions from air pollution sources. 2. C1 through C30 organic compounds from medium duty diesel trucks. Environ. Sci. Technol. 33, 1578–1587.
Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, B.R.T., 2002. Measurement of emissions from air pollution sources. 5. C1-C32 organic compounds from gasoline-powered motor vehicles. Environ. Sci. Technol. 36, 1169–1180.
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, 2097–2098.
Shim, W.J., Hong, S.H., Yim, U.H., Kim, N.S., Oh, J.R., 2002. Horizontal and vertical distribution of butyltin compounds in sediments from shipyards in Korea. Arch. Environ. Contam. Toxicol. 43, 277–283.
Smith, B.L., Bruno, T.J., 2007. Composition-explicit distillation curves of aviation fuel JP-8 and a coal-based jet fuel. Energy Fuel 21, 2853–2862.
Soman, V., Suresh, A., Remani, A.K., Venugopalan, V.K., Keedakkadan, H.R., 2024. A holistic review on diverse lipid biomarkers as environmental indicators: extraction and analysis from sediments to microbial communities. Geomicrobiol J. 41, 891–909.
Stratton, C.A., Thompson, Y., Zio, K., Morrison III, W.R., Murrell, E.G., 2024. uafR: An R package that automates mass spectrometry data processing. PLoS One 19, e0306202.
Szymańska, E., Saccenti, E., Smilde, A.K., Westerhuis, J.A., 2012. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 8, 3–16.
Tate, S., Watanabe, Y., Chiba, A., 1993. Synthesis of ultra-high molecular weight poly (ethylene terephthalate) by swollen-state polymerization. Polymer 34, 4974–4977.
Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and nontarget screening for contaminants of emerging concern in an urban estuary. Environ. Sci. Technol. 54, 889–901.
Ting, Y.F., Praveena, S.M., 2017. Sources, mechanisms, and fate of steroid estrogens in wastewater treatment plants: a mini review. Environ. Monit. Assess. 189, 178.
Ulrich, N., Schüürmann, G., Brack, W., 2013. Prediction of gas chromatographic retention indices as classifier in non-target analysis of environmental samples. J. Chromatogr. A 1285, 139–147.
Wakoli, J., Anjum, A., Sajed, T., Oler, E., Wang, F., Gautam, V., LeVatte, M., Wishart, D. S., 2024. GCMS-ID: a webserver for identifying compounds from gas chromatography mass spectrometry experiments. Nucleic Acids Res. 52, 381–389.
Wang, J., Gong, X., Hu, M., Zhao, L., 2023. Improved GSimp: a flexible missing value imputation method to support regulatory bioequivalence assessment. Ann. Biomed. Eng. 51, 163–173.
Wang, X., Yu, N., Yang, J., Jin, L., Guo, H., Shi, W., Zhang, X., Yang, L., Yu, H., Wei, S., 2020. Suspect and non-target screening of pesticides and pharmaceuticals transformation products in wastewater using QTOF-MS. Environ. Int. 137, 105599.
Wei, R., Wang, J., Jia, E., Chen, T., Ni, Y., Jia, W., 2018a. GSimp: a Gibbs sampler based left-censored missing value imputation approach for metabolomics studies. PLoS Comput. Biol. 14, e1005973.
Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., Ni, Y., 2018b. Missing value imputation approach for mass spectrometry-based metabolomics data. Sci. Rep. 8, 663.
Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2023. Dplyr: a grammar of data manipulation. R package version 1 (1), 4. Available online: https://CRAN.R-project.org/package=dplyr/.

Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. The CompTox chemistry dashboard: a community data resource for environmental chemistry. J. Chemother. 9, 61.

Woods, R.W., Letinski, D.J., Febbo, E.J., Dzamba, C.L., Connelly, M.J., Parkerton, T.F., 2007. Assessing the aquatic hazard of commercial hydrocarbon resins. Ecotoxicol. Environ. Saf. 66, 159–168.

Xia, J., Wishart, D.S., 2011. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. Curr. Protoc. Bioinformatics 34, 14–10.

Xie, Y., Cheng, J., Tan, X., 2024. DT: a wrapper of the JavaScript library 'DataTables'. R package version 0, 33. Available online: https://CRAN.R-project.org/package=DT/.

Yang, J., Zhao, X., Lu, X., Lin, X., Xu, G., 2015. A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. Front. Mol. Biosci. 2, 4.

Yokoyama, H., Choi, J.-W., 2010. New records of three *Paraprionospio* species (Polychaeta: Spionidae) from Korean waters. Ocean Sci. J. 45, 55–61.

Zhang, M.-Y., Su, Q.-Z., Lin, Q.-B., Zhong, H.-N., Chen, S., Li, D., Dong, B., Xu, M.-Q., Li, W.-L., 2025. Identification and quantification of volatile organic compounds in Chinese recycled HDPE: consequences for its applicability. Food Packag. Shelf Life 47, 101424.

Zheng, W., Wang, X., Tian, D., Jiang, S., Andersen, M.E., He, G., Crabbe, M.J.C., Zheng, Y., Zhong, Y., Qu, W., 2013. Water pollutant fingerprinting tracks recent industrial transfer from coastal to inland China: a case study. Sci. Rep. 3, 1031.