Environmental Science & Technology

pubs.acs.org/est

Article

Application of Nontarget High-Resolution Mass Spectrometry Fingerprints for Qualitative and Quantitative Source Apportionment: A Real Case Study

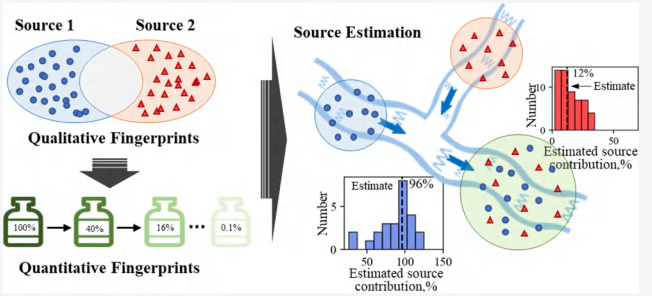
Di Xia, Lijun Liu, Bo Zhao, Danping Xie, Guining Lu, and Rui Wang*

Cite This: https://doi.org/10.1021/acs.est.3c06688

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: High-resolution mass spectrometry (HRMS) provides extensive chemical data, facilitating the differentiation and quantification of contaminants of emerging concerns (CECs) in aquatic environments. This study utilizes liquid chromatography-HRMS for source apportionment in Chebei Stream, an urban water stream in Guangzhou, South China. Initially, 254 features were identified as potential CECs by the nontarget screening (NTS) method. We then established 1689, 1317, and 15,759 source-specific HRMS fingerprints for three distinct sources, the mainstream (C3), the tributary (T2), and the rain runoff (R1), qualitatively assessing the contribution from each source downstream. Subsequently, 32, 55, and 3142 quantitative fingerprints were isolated for sites C3, T2, and R1, respectively, employing dilution curve screening for source attribution. The final contribution estimates downstream from sites C3, T2, and R1 span 32−96, 12−23, and 8−23%, respectively. Cumulative contributions from these sources accurately mirrored actual conditions, fluctuating between 103 and 114% across C6 to C8 sites. Yet, with further tributary integration, the overall source contribution dipped to 52%. The findings from this research present a pioneering instance of applying HRMS fingerprints for qualitative and quantitative source tracking in real-world scenarios, which empowers the development of more effective strategies for environmental protection.



KEYWORDS: *nontarget screening, HRMS fingerprints, qualitative analysis, dilution curves, quantitative source apportionment*

INTRODUCTION

With the increasing production and use of synthetic organic substances in the domestic, agricultural, and industrial sectors, contaminants of emerging concerns (CECs) (e.g., pharmaceuticals, pesticides, personal care products, etc.) are frequently detected in urban waters, which often adversely impacts both humans and ecosystems.[1−3] However, differentiating and quantifying the sources of CECs in aquatic environments remains a challenge due to the complex nature of the water systems and the presence of multiple sources of contamination.[4−6] Previous target methods for source tracing have predominantly depended on well-established indicator compounds such as polycyclic aromatic hydrocarbons,[7] pesticides[8] and metals,[9] and more, utilizing analytical techniques like gas chromatography−mass spectrometry (GC−MS) and liquid chromatography−mass spectrometry (LC−MS), among others. Although with high sensitivity and specificity for the contaminants selected, these approaches inevitably miss other contaminants that are not included in the target list and be constrained by the lack of reference standards for CECs.[10] Therefore, a comprehensive analysis is critical for accurately characterizing their occurrence and sources.[11,12]

High-resolution mass spectrometry (HRMS) is esteemed for its ability to indiscriminately detect ionizable chemicals with high mass accuracy, making it unparalleled in analytical chemistry. This capability has become increasingly prominent in recent years, particularly in the field of screening CECs.[13−15] Coupled with chromatography, HRMS can detect hundreds to thousands of chemicals, providing a more exhaustive picture of contaminants present in the samples.[16−19] The comprehensive data derived from nontarget screening (NTS) allows for the creation of unique chemical fingerprints, presenting a significant opportunity to elucidate the complex chemical composition of samples and trace the origins of pollutants.[13,20−22] Although such methodologies have been successfully employed in food and herbal medicine authentication to differentiate and authenticate samples,[23,24] their application in the environmental sector is less widespread.[25] Dávila-Santiago used machine learning tools for chemical fingerprinting and environmental source tracking, which

Received: August 16, 2023
Revised: November 30, 2023
Accepted: December 1, 2023

required extensive training samples, and were constrained by varying characteristics of the same pollution source (such as wastewater treatment plants) across different regions, as well as intrinsic uncertainties within the methods themselves.[26] Peter et al. employed dilution curves with hierarchical cluster analysis (HCA) for the selection of fingerprint compounds to differentiate and quantify pollutant source contributions from disparate samples of specific source types and assessed the reliability in increasingly complex background matrices.[27,28] However, this proven approach has been primarily utilized in controlled laboratory environments without being extended to the field.

Given this groundwork, the current study aims to extend the application of the fingerprint strategy beyond the laboratory setting. We aspire to adapt this methodology for real-world applications, acknowledging the complexities and variations inherent in these environments. We remain cognizant of the unique challenges posed by such environments, yet we also recognize the vast potential this method holds for improving source estimation in diverse contexts. In this study, the Chebei stream was chosen as a representative urban water with multiple inputs (such as tributaries, rain runoff, etc.) in Guangzhou, South China. The objectives were to (1) characterize and trace potential source-specific CECs using the NTS method; (2) differentiate the nontarget fingerprints and qualitatively track their distribution in Chebei stream; and (3) isolate the quantitative fingerprints and estimate the source contribution. This research aspires to harness the vast potential of HRMS for improving source estimation in diverse real-world contexts.

MATERIALS AND METHODS

Chemicals. The analytical standards were purchased from Alta Scientific (Tianjin, China). Formic acid of LC/MS-grade with purity exceeding 99% was procured from DiKMA Technologies, California. LC/MS-grade ammonium acetate, with purity greater than 99%, was sourced from CNW Technologies, Shanghai, China. Methanol (both LC/MS and high-performance liquid chromatography (HPLC)-grade), LC/MS-grade water, and a variety of solvents, including dichloromethane, acetone, and *n*-hexane (all HPLC-grade), were obtained from Fisher Scientific, Pittsburgh, PA.

Sample Collection, Preparation, and Extraction. Water samples were collected on May 16, 2022, following a rainfall, from selected sites along the Chebei stream (sites C1−C9, as shown in Figure S1). Within this stream, a tributary was identified as a presumptive source of contamination from which samples were gathered from sites T1 and T2. Another source sample was collected from site R1, which represents a potential source of rain runoff and is located in the proximity of site C6. We utilized precleaned 4 L amber glass bottles for sample collection, which were thoroughly rinsed with hexane, acetone, and methanol prior to use. All samples were stored at 4 °C in a refrigerator and were extracted within a period of 24 to 72 h. To isolate the quantitative fingerprint for each source, we assessed the peak area responses using source dilution curves of all nontarget detections found in the source fingerprints of three distinct sources: C3 represents the mainstream source, T2 represents the tributary source, and R1 represents the rain runoff source. This was performed across various source concentrations (100, 40, 16, 6, 2.5, 1, and 0.1% v/v). The term "fingerprint" refers to the collective nontarget HRMS detection characteristic of a specific sample or source. It is important to note that dilution was conducted prior to solid phase extraction (SPE) by means of dilution with LC/MS-grade water.

From the initial 4 L of water, 3 L was divided into three 1 L bottles for triplicates. Each water sample was initially filtered utilizing 0.45 μm glass fiber membranes (50 mm diameter, Jinteng, China) and subsequently spiked with a set of 10 isotope-labeled internal standards (Table S1). Replicated extractions of the samples were conducted using an automated SPE instrument (Fotector Plus, RayKol, China) in conjunction with an Oasis HLB cartridge (6 cc, 500 mg, Waters). Prior to extraction, the Oasis HLB cartridge underwent preconditioning with 10 mL of methanol and 10 mL of deionized water at a consistent flow rate of 3 mL/min. Water samples (1 L) were then introduced to the preconditioned SPE cartridges and maintained at a flow rate of 12 mL/min. Following this step, the cartridges were rinsed with deionized water (10 mL, 4 mL/min), air-dried under nitrogen for 25 min, and then eluted with methanol (administered twice, 5 mL each time). The eluates (100% methanol) were subsequently reduced to 1 mL (the enrichment factor was 1000-fold) by nitrogen, employing an auto multiple sample concentrator-MPEva GS (Relabor Instruments, Guangzhou, China), and then spiked with five isotopically labeled internal standards (Table S1). To conclude, the extracts were transferred to 2 mL amber glass vials and stored at −20 °C until further analysis was conducted.

Analytical Methods. Extracts analysis followed methods described previously.[29,30] Briefly, the extracts underwent analysis through ultraperformance liquid chromatography (UPLC), coupled with an Xevo G2-XS quadrupole time-of-flight HRMS (Waters). The analysis was conducted in two separate runs, one using electrospray ionization (ESI) in positive mode (ESI+) and the other in negative mode (ESI−). UPLC separation was achieved by employing an ACQUITY UPLC BEH C18 column (150 mm × 3.0 mm, 1.7 μm, Waters) complemented by a guard column (2.1 mm × 5 mm, 1.7 μm), both operating at a temperature of 40 °C. The system was set to introduce an injection volume of 1 μL at a flow rate of 0.35 mL/min. For ESI+, the mobile phase was composed of 5 mM ammonium acetate and 0.1% formic acid in water (component A), and pure methanol (component B). For ESI−, the mobile phase consisted of 0.05% formic acid in water (A) and methanol (B), respectively. The binary gradient program was set as follows: 2% B from 0 to 0.5 min, 98% B from 18 to 24 min, reverting to 2% B at 24.1 min; with a total runtime of 27 min. For mass spectrometric analysis, the extracts were subjected to independent analyses by both data-independent acquisition (DIA) and data-dependent acquisition (DDA). For DIA, the MS^E mode was adopted to capture both MS^1 and MS^2 data in the m/z range of 50−1000. The MS^1 spectra were acquired with a collision energy of 0 eV, whereas the MS^2 spectra were generated using a ramping collision energy between 15 and 45 eV, with a scan time set at 0.2 s/spectrum. For DDA, MS^2 data were gathered for the seven most abundant masses using the same m/z range (50−1000), collision energy (15−45 eV), and total scan rate (5 Hz) with DIA mode.

To ensure quality assurance and control (QA/QC), mass accuracy was regularly calibrated by injecting a Leucine Enkephalin solution every 15 s before each analytical run. The mass resolution for the instrument, as measured using Leucine Enkephalin with m/z of 556.2771 in positive mode

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

and 554.2615 in negative mode (used as the lock mass for the Waters Xevo G2-XS quadrupole time-of-flight HRMS), stood at approximately 30,000. Triplicates of laboratory and field blanks were extracted and processed using the same materials and methods used for the samples. For every 12 samples, two solvent blanks (pure methanol) and one internal standard (ISTD) control sample, incorporating 10 extracted and five instrumental ISTDs, were sequentially analyzed to monitor the analytical performance. Deviations of ISTD retention time (RT) were maintained below 0.1 min, and mass errors of ISTDs did not exceed 10 ppm in all of the samples.

Data Reduction and Analysis. We implemented a self-developed Python package, PyHRMS (https://pypi.org/project/pyhrms/), which has been utilized in previous works,[29−33] to extract and align compound features (characterized by RT-exact mass pairs), group isotopes/adducts into nontarget compounds, filter detections, and perform statistical analyses. Prior to this analysis, the raw data was transformed into the mzML format via MSconvert.[34] Only signals with a mass intensity >500, peak areas >5-fold relative to blanks (including solvent and field blanks), and p value <0.05 (determined through p-tests comparing each set of sample data to solvent and field blanks) were selected for further analysis. Features originating from various adducts, such as $[M + H]^+$, $[M + Na]^+$, $[M + NH_4]^+$, $[M + K]^+$ adducts, and C/Cl/S/Br isotopologues, were identified and grouped together as a single component. MS^2 spectrum information, procured from either DIA or DDA data, was assigned to each distinct feature. Subsequently, these features underwent matching with three principal databases: (1) an in-house database encompassing approximately 2000 CECs, which include pharmaceuticals, personal care products, pesticides, and veterinary drugs, among others, with each compound preanalyzed to register its corresponding RT, m/z, and MS^2 fragments; (2) NORMAN Suspect List Exchange (NORMAN-SLE);[35] (3) Massbank of North America (MoNA) (https://massbank.us/downloads, LC-MS Spectra).[36]

Compound identifications were attributed to confidence levels based on their degree of alignment with reference standards and MS^2 libraries.[37] The highest level of confidence (Level 1) was conferred upon matches with accurate mass (deviation <10 ppm), RT (deviation <0.1 min), and at least two matching MS^2 fragment (deviation <0.015 Da) with reference standards. Level 2 confidence was assigned to matches with accurate mass and a minimum of three fragment ions matching the MS^2 libraries. The graphical representations elucidating the identification confidence levels (levels 1 and 2) are depicted in Figure S2. For compounds matched with RT (deviation <0.1 min), MS^1 (deviation <10 ppm), and with no more than one corresponding MS^2 fragment (deviation <0.015 Da) and compounds only matched with accurate mass and no more than two fragment ions, they were classified as level 4. In the methodology of our study, we focused on the quantitative match of MS^2 fragments, choosing not to consider their relative abundance. Although the direct comparison of raw spectral data is theoretically optimal, it relies on the premise that database spectra and measured spectra are produced under the same collision energy conditions—a scenario that is frequently unattainable with the use of online databases. The consideration of fragment abundance in the comparison process can inadvertently result in false negatives. Our chosen approach, which emphasizes the detection of the fragment presence, is designed to circumvent this issue. This method has been previously validated in our earlier research,[38] reinforcing its reliability and effectiveness in MS^2 comparison under variable analytical conditions.

Detailed information on all identifications is available in Table S2. Subsequently, compounds matched with RT and MS^1 underwent semiquantification via the relative response factor (RRF) method,[30] employing a five-point internal standard calibration curve and extraction ISTD (i.e., Atrazine-D5 or Propylparaben-D4). The RF of the target compound was initially calculated by using its peak area and known concentration. Then, the RRF, calculated by comparing the RF of the target compound to the RF of the reference standard, was applied to determine the concentration of the target compound in the sample.

HCA was conducted utilizing seaborn[39] and ComplexHeatmap[40] packages in python, to compare and differentiate the chemical profiles of the samples. The HCA primarily utilized Euclidean distances as a measure of dissimilarity among the data points, and the linkage criteria applied during the analysis defaulted to the 'single' method. All other parameters in the analysis adhered to their default settings.

RESULTS AND DISCUSSION

Tracing Identified CEC Origins in Chebei Stream. A comprehensive analysis of all samples from Chebei Stream resulted in the recognition of 26860 distinct features across both ionization modes. After an extensive database comparison, we identified 254 features with different degrees of confidence. To detail, 45 features corresponded to level 1 (matching RT, MS^1, and a minimum of 2 fragments), 77 to level 2 (matching MS^1 and at least 3 fragments), and 132 to level 4 (either matching RT, MS^1, and up to 1 fragment, or just MS^1 and up to 2 fragments) (Table S2). The identified compounds encompassed a diverse range of categories, including pharmaceuticals (n = 86), pesticides (n = 70), industrial materials (n = 39), natural products (n = 22), veterinary drugs (n = 14), personal care products (n = 10), plasticizers or flame retardants (n = 8), and food additives (n = 5). The considerable presence of both pharmaceuticals and pesticides can be ascribed to the pathway of the stream through densely populated regions and its proximity to extensive farmlands. Urban areas contribute pharmaceutical compounds, stemming from domestic use, into the sewage system.[41] Farmlands, on the other hand, introduce pesticides, which are prevalent in modern agriculture, into the surrounding environment.[42] During periods of rainfall, runoff from these areas, which may contain both agricultural chemicals and, in some cases, domestic wastewater, enters the stream. This leads to the heightened occurrence of these compounds.[43]

The distribution and prevalence of identified CECs at each site (Figure S3) underscore the stark compositional disparity between the mainstream and its intersecting tributaries (sites T1 and T2) as well as the rain runoff (site R1). Initial sampling at site C1 yielded 112 detected CECs, increasing to 160 at the final site, C9. Notably, while 50 of the identified CECs were universally detected across all sites, 40 were exclusively present at either the tributaries or rain runoff. The influx from tributaries and the rain runoff shifts the mainstream's composition, thereby accentuating the pivotal role these water sources play. For the objectives of this research, although these compounds may not necessarily originate from

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

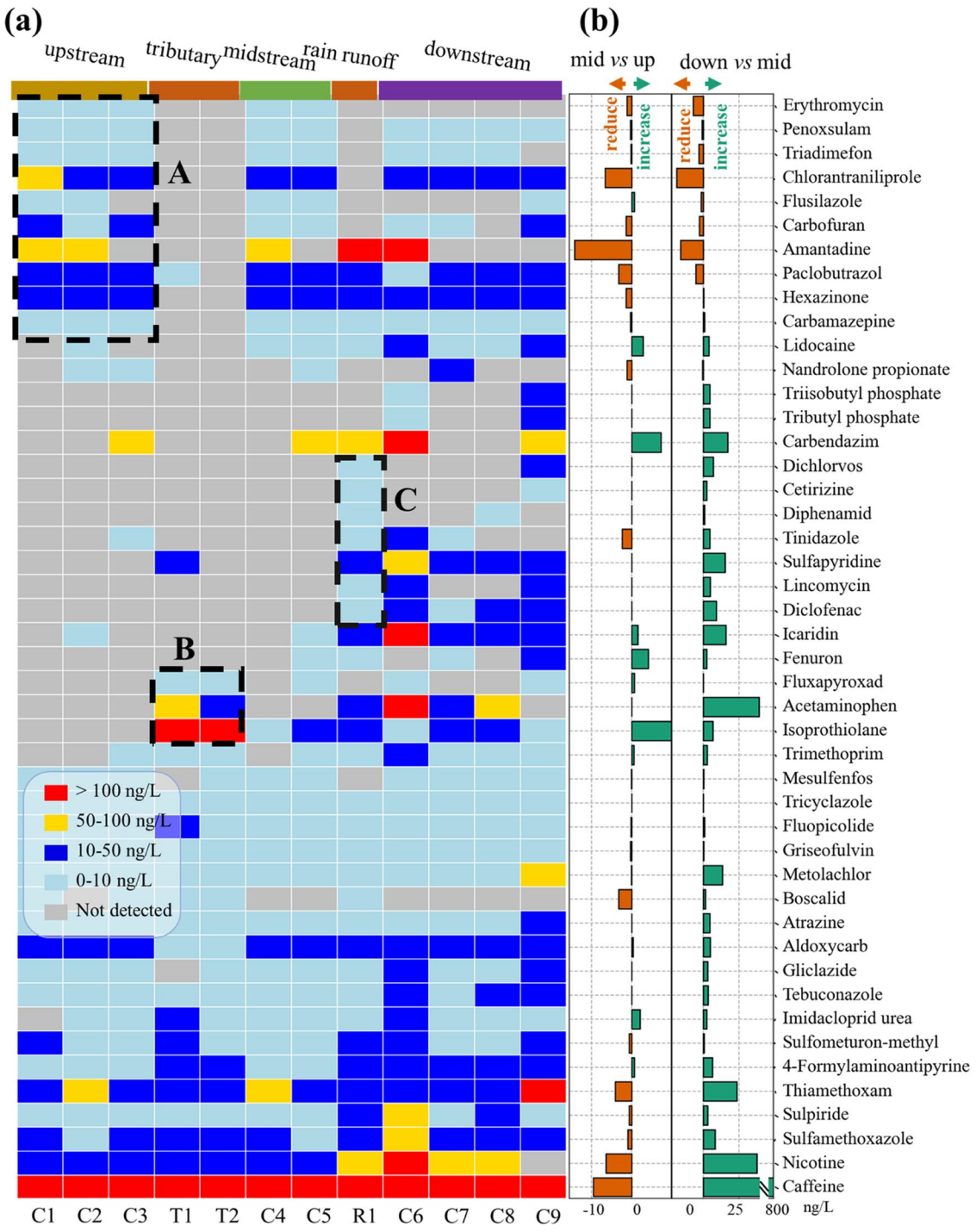


Figure 1. (a) Heatmap for the concentration level of 46 CECs (detected at a minimum of two points) in sampling sites (the concentration is available in Table S3). Boxes A, B, and C highlight groups of indicator compounds that were origin from upstream, tributary, and rain runoff, respectively. (b) Average pollutant concentrations change between midstream and upstream (mid vs up), downstream and midstream (down vs mid) areas. The carrot and green represent a decrease and increase of concentrations, respectively.

conspicuous point sources, we have opted to consider the tributaries and rain runoff as point source contamination sites.

In our pursuit to deepen our understanding of CECs distribution, we focused our analysis on certain compounds for which we have established standards. As shown in Figure 1, the semiquantitative data reveals that these 46 CECs (Table S3), detected at two or more sites, displayed average concentrations ranging from 0.30 to 1111.4 ng/L, and median concentrations from 0.0 to 492.14 ng/L. The introduction of water from the tributaries and rain runoff instigated fluctuations in compound concentrations. Specifically, the tributary input induced an increase in concentration for 14 compounds and a decrease in concentration for 23. In parallel, the rain runoff input incited an increase for 37 compounds and a decrease for 9. These variations in compound concentrations following the tributary and rain runoff inflows indicate two key processes. The compounds are either introduced by these sources (i.e.,

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

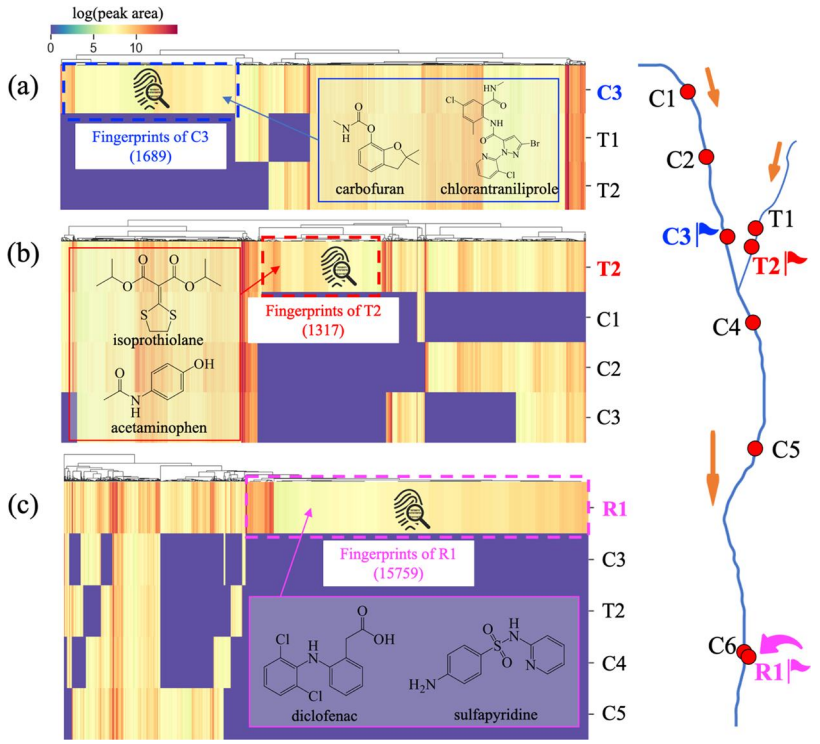


Figure 2. Hierarchical cluster analysis of the (a) C3 and tributary source type (i.e., T1 and T2), (b) T2 and upstream source type (i.e., C1, C2, and C3), and (c) R1 and upstream source type (i.e., C3−C5 and T2). Each row represents an average of three replicate samples, and each vertical line represents an individual compound with color representing log(peak area) (absent = dark blue, increasing to yellow, light red, and dark red with increasing peak area). The dotted box highlights groups of indicator compounds that were completely unique to the other source type. The solid box shows representative identified unique CECs.

tributaries or rain runoff), or their concentrations are affected due to dilution by the incoming water.

Our comparative examination of CEC concentrations in the mainstream and its tributaries has provided key insights into the potential origins of specific contaminants. A suite of 10 pollutants, including erythromycin, penoxsulam, triadimefon, etc., which had been frequently detected in surface waters,[44−46] were discernible both upstream and downstream but were notably absent in the tributary. This observation suggests that the sources of these pollutants are likely located upstream. Conversely, fluxapyroxad, acetaminophen, and isoprothiolane, undetected upstream, were found at both the tributary T2 and the downstream locations. These findings could potentially implicate tributary T2 as a source of these pollutants. This highlights the site-specific distribution of these compounds, even within the homogeneous urban context encompassing both the mainstream and its tributaries. In addition, dichlorvos, cetirizine, diphenamid, etc. showed a pronounced presence downstream of R1, with minimal detection at site C5 and upstream. This implies that R1 could be a plausible source of these contaminants. These substances, commonly found in agricultural, residential, and medical settings, can infiltrate water systems via rain runoff, improper disposal practices, or leakage from sewage infrastructures.[47−49] While previous studies do not specifically detail the occurrence of these substances, the general understanding of their sources and pathways supports our findings and interpretations.

We are currently providing a preliminary understanding of the contribution of distinct sources to specific contaminants. However, the full range of contaminants, particularly those not yet identified, is not entirely accounted for in our current analysis. In the downstream regions, where the confluence of multiple sources occurs, the attribution of contamination to a particular source becomes a complex endeavor. This complexity is amplified when our analysis is constrained exclusively to identifiable end-members exclusively. The dynamic contributions from various sources and the origins of additional unidentified compounds thus represent key areas necessitating comprehensive investigation in our further analysis.

Isolation and Application of Fingerprints for Qualitative Analysis. Tributaries flowing into a river can significantly alter the presence and composition of CECs within the water body. Such inflow may dilute specific CECs and introduce novel contaminants.[50,51] To thoroughly understand the impacts of incoming water on the mainstream, we applied a fingerprinting strategy based on HRMS. HRMS fingerprints from each source usually relate to particular environmental phenomena, providing an opportunity for qualitative source differentiation before engaging in extensive identification efforts.[52−54] It is crucial to note that successful source tracking hinges on the ability to establish compositional uniqueness. Sources with minimal chemical complexity might present differentiation and tracking challenges.[27]

We utilized hierarchical cluster analysis (HCA) to establish source-specific HRMS fingerprints by identifying unique compounds for each source. As depicted in Figure 2a, we discovered that 67% (n = 3361) of 5050 nontarget compounds in site C3 were also detected in one or both tributary sites (i.e., T1 and T2). This overlap is attributed to the shared water origin (i.e., rain runoff or the upstream Pearl River), leading to similar primary compositions in the mainstream and tributaries. However, 33% (n = 1689) of the detected C3

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

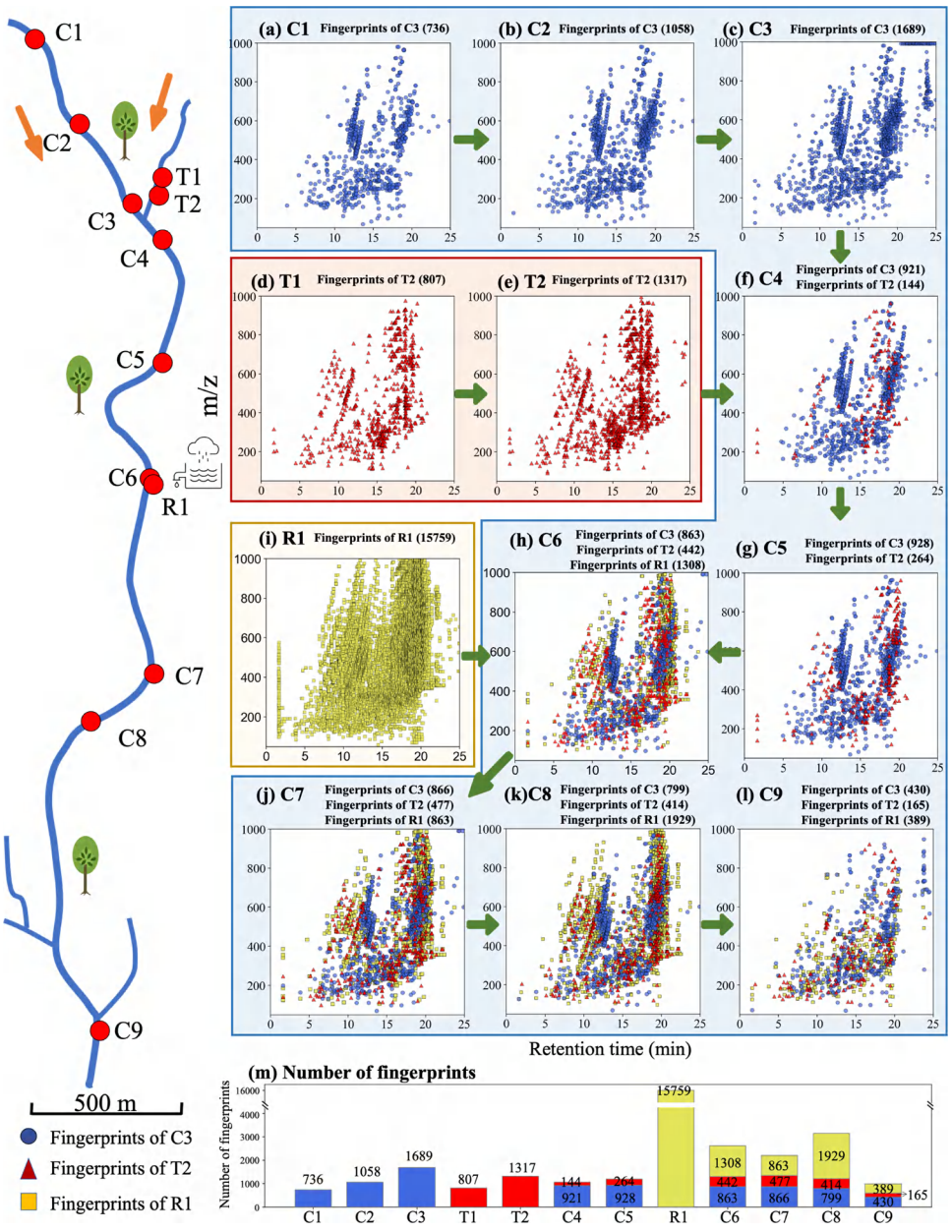


Figure 3. (a–l) Distribution of C3, T2, and R1 fingerprints in each site. Blue points, red triangles, and yellow squares represent C3, T2, and R1 fingerprints, respectively. (m) The number of C3, T2, and R1 fingerprints detected in each site.

signature compounds were exclusively distinct and absent in tributary source types. These unique C3 compounds, including 26 identified CECs such as carbofuran, chlorantraniliprole, penoxsulam, triadimefon, etc., corroborated our semiquantification results. The unique C3 composition compared to those of T1 and T2 can be attributed to the disparate paths of the mainstream and tributaries, which inevitably introduce diverse compounds. These unique compound compositions form a distinct fingerprint for each source, paving the way for further applications.

Similarly, 24% (n = 1317) of the 5527 T2 compounds were completely unique to the upstream source type (C1–C3) (Figure 2b). This group included 12 identified CECs, such as isoprothiolane, acetaminophen, etc. Meanwhile, 65% (n = 15,759) of R1 signature compounds were entirely unique to the upstream source type (C3–C5 and T2) (Figure 2c). This group included 40 identified CECs, such as diclofenac, sulfapyridine, lincomycin, etc. These fingerprint compounds (Table S4) deliver more statistical power than target analysis for tracking and differentiating sources, emphasizing their significance in the study of environmental contaminants.

Upon establishing the unique fingerprints for each source, we can effectively trace these signatures from upstream to downstream sites. Illustrated in Figure 3a–c, the number of C3 fingerprint compounds in upstream sites (i.e., C1 and C2) is 736 and 1058, respectively. When the tributary water merges with the mainstream, it introduces T2 source fingerprint compounds (Figure 3e). The water at site C4, now a confluence of the two sources, contains 921 and 144 fingerprint compounds from the C3 and T2 sources, respectively. The observed decrease in the number of fingerprint compounds can be attributed to three primary reasons: (1) dilution from both sources leading to lower detection limits for some compounds, particularly those with

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX–XXX

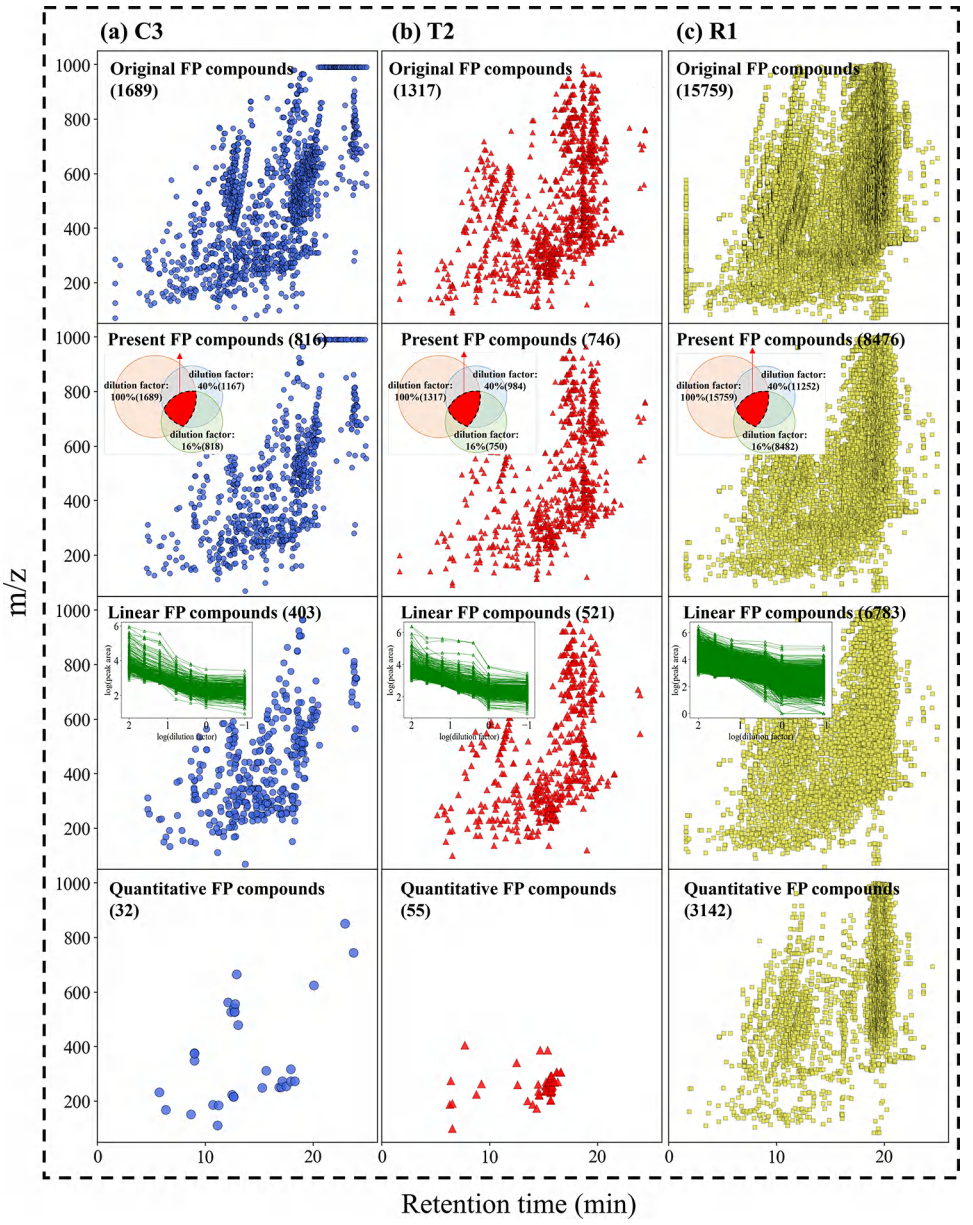


Figure 4. Screening process of C3, T2, and R1 quantitative source fingerprints, respectively. Original fingerprint (FP) compounds represent the fingerprint compounds established for qualitative analysis. Present FP compounds represent the fingerprint compounds presented in at least the top three highest dilution levels (100, 40, and 16%). Linear FP compounds represent the fingerprint compounds that also met dilution curve linearity requirements (peak area decreased with increasing dilution, $R^2 \geq 0.8$, slope ≥ 0.3). Quantitative FP compounds represent the compounds with peak areas at least five times greater than the extracted peak area of similar m/z values ($\pm$0.015 Da) within comparable retention time (RT) ranges ($\pm$0.2 min).

low intensities; (2) uneven compound distribution due to incomplete mixing of water from the mainstream and the tributaries; and (3) potential sorption or degradation of compounds as the merged water flows toward site C4.

As the water progresses downstream, the number of fingerprint compounds from C3 remains relatively stable (ranging from 799 to 928), while the number of fingerprint compounds from T2 increases from 144 at site C4 to 477 at site C7 and 414 at site C8. This increase suggests that water from the tributaries is progressively mixing with the mainstream. This mixing effect was also noted when rain runoff, characterized by a substantial 15,759 fingerprint compounds at source R1, entered the mainstream, where it decreased to 1308 at site C6 and 863 at site C7 and then rose to 1929 at site C8. Nevertheless, at site C9, the number of fingerprint compounds from all three sources significantly decreased. The reduction (from 799, 414, and 1929 to 430, 165, and 389 for C3, T2, and R1 sources, respectively) occurs due to the influx of two additional tributaries (Figure S1), which considerably dilute all the fingerprint compounds. We note that, despite site C9 being over 4 km away from sources C3 and T2 and 3 km away from R1—with the presence of other tributary and rain runoff inputs into the mainstream—we can still discern a portion of the original fingerprints. This robust tracking ability underscores the effectiveness of the fingerprinting technique, akin to isotopic tracing, in pinpointing sources even in the face of dilution and other complex environmental factors.

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

The distribution of source fingerprint features in C3, T2, and R1 offers a qualitative evaluation of fingerprint fidelity in the midst of reduced source concentrations and a complex background matrix. The focus on occurrence, without accounting for relative abundance, however, suggests that further assessments are needed for the more quantitative applications of these fingerprints. This fingerprinting approach, which parallels isotopic tracing, is not only efficient but also a cost-effective solution for tracing contaminants back to their sources.

Criteria for Quantitative Fingerprint Screening and Its Application in Source Estimation. Quantitative source estimation, as illuminated by early studies,[27,28,55] has shown promising results particularly when employing dilution curves for quantitative fingerprint compound selection. In order to understand the dilution behavior and select appropriate fingerprint compounds, we conducted dilution curves for samples from C3, T2, and R1 sources. The process of sample dilution inevitably affects the detectability of certain features, pushing some below the detection thresholds. The low-intensity features, despite their potential significance, may not be suitable for quantitative analysis due to matrix dilution effects.[56,57] However, our study circumvented the consideration of environmental matrices, instead utilizing source sample dilution curves and raw peak area data, given the absence of established standards for nontarget data.[27] Previous studies support this approach, suggesting that normalizing to a single internal standard does not consistently improve precision and could potentially yield inferior results.[58,59]

To facilitate precise and quantitative estimation of the source contribution, we employed a series of filtering criteria, consistent with the methods used by Peter et al.[27] These criteria were designed to isolate nontarget compounds that remained consistent during the dilution process (Figure 4). Upon the unique fingerprints established for qualitative analysis (1689, 1317, and 15,759 of C3, T2, and R1, respectively), compounds found in at least the top three highest dilution levels (100, 40, and 16%) were remained. This filter criteria reduced the number of nontarget compounds in the C3, T2, and R1 to 816, 746, and 8476 compounds, respectively. Subsequently, the remaining compounds were subject to dilution regression analysis. In particular, peak area data were logarithmically transformed, and a dilution regression was computed for each nontarget compound. Nontarget compounds were retained if their dilution curve demonstrated a peak area decrease as the source concentration diluted and had $R^2 \geq 0.80$ (minimum linearity), a slope ≥ 0.30 (equivalent to $\geq$2-fold peak area change for every 10-fold concentration change), and p value <0.05. These criteria narrowed down the number of compounds, resulting in 403, 521, and 6783 compounds in the C3, T2, and R1, respectively.

In a real-world context, the application of HRMS fingerprints for quantitative analysis encounters unique challenges. Here, target compounds are diluted by complex natural waters, leading to a nonproportional peak area reflection of the dilution factor. This variability in peak detection could be attributed to potential interference from chromatographic peaks that have close m/z and RT, or from high background signals present in the dilution water. To address these concerns, our final filtering criterion retained compounds whose peak areas were at least five times greater than the extracted peak area of similar m/z ($\pm$0.015 Da) within similar RT ranges ($\pm$0.2 min). This measure reduces the impact of nonunique chromatographic features from other sources, allowing the focus to remain on unique, distinguishable compounds. Consequently, the number of nontarget compounds in the C3, T2, and R1 further reduced to 32, 55, and 3142 respectively. These remaining unique features in each source were then utilized for subsequent quantification processes, thereby improving source estimation precision by reducing potential confounding factors.

For the remaining 3229 quantitative source fingerprints (32, 55, and 3142 in C3, T2, and R1, respectively), an individual source contribution was calculated for each compound based on the observed peak area, indicative of the extent of source dilution. This procedure is comparable to employing a specific chemical calibration curve to determine an unknown concentration, thus, generating 3229 unique source contribution estimations for each mixture. By leveraging these quantitative data distributions, we were able to infer the final source contribution, portrayed as the median of these multiple individual estimates for that some of these unidentified compounds may trend toward over- or underestimation, while others may produce accurate values. While peak area is not a flawless surrogate for chemical concentration, the peak area responses consistent with source concentration are amenable to targeted, quantitative approaches, or site modeling efforts. This approach emphasizes the robustness of the analysis and underscores the importance of cautious interpretation when dealing with data derived from nontarget analysis.[27]

The final contribution estimates based on the median values derived from multiple individual source contribution estimations for C3, T2, and R1 are listed in Figure 5. By evaluating their flow rates and cross-sectional areas, we derive an estimate of approximately 82 and 18% for the flow contributions from the mainstream and tributary, respectively. However, we have no means to estimate the flow quantity of the rain runoff. The estimated flow quantities of the mainstream and tributary were employed to cross-verify the source contributions determined through our fingerprinting strategy. At site C4, the mainstream source (i.e., from site C3) accounted for a 96% contribution, while the tributary (i.e., from site T2) contributed approximately 12%. These contribution rates align closely with the estimated flow quantities (mainstream: 82%, tributary: 18%), and the total contribution summed to 108%, approximating the ideal 100%. Likewise, at site C5, the contributions from the mainstream and tributary were 95 and 18%, respectively, resulting in a combined contribution of 113%. The increase in tributary contribution from 12% at site C4 to 23% at site C7 is mainly attributable to the mixing process during flow. This consistency is reflected in the qualitative analysis (Figure 3), where the count of fingerprint compounds from tributaries increased from 144 at site C4 to 477 at site C7.

When rain runoff is introduced to the mainstream at C6, the contributions from the mainstream, tributary, and rain runoff (i.e., site R1) are 73, 22, and 19% respectively. The cumulative contribution of all three sources ranged between 103 and 114% from C6 to C8, closely aligning with the actual conditions. The contribution of R1 fluctuated from 19% at site C6 to 17% at site C7, before it escalated to 23% at site C8. This variation aligns with the qualitative analysis (Figure 3), which revealed a decrease in the count of rain runoff fingerprints from site R1, from 1308 at site C6 to 863 at site C7, before increasing to 1929 at site C8. This fluctuation is primarily attributed to uneven mixing along the mainstream. Finally, at site C9, the

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

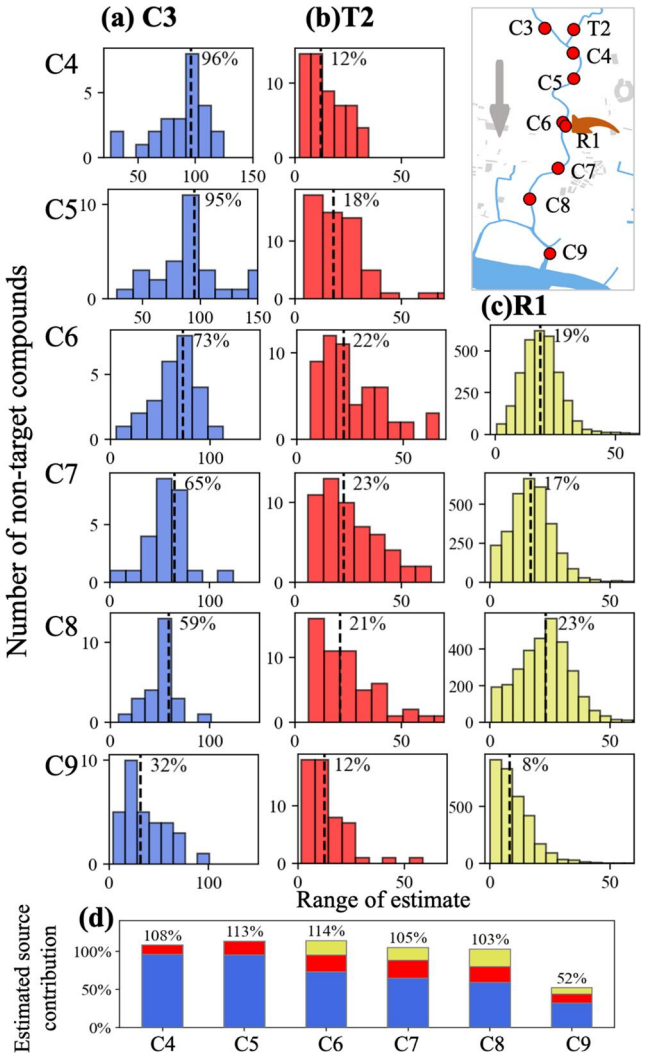


Figure 5. Histograms showing the distribution of (a) C3, (b) T2, and (c) R1 contribution estimates and number of contributing nontarget compounds for the mixtures from sites C4 to C9. Estimates using the 3229 quantitative source fingerprints (32, 55, and 3142 in C3, T2, and R1, respectively). The dashed line and percentage number indicate the median estimated source contribution. (d) The total estimated source contribution of different sources from C4 to C9. Blue, red, and yellow bars represented C3, T2, and R1 source contribution estimates, respectively.

introduction of two additional tributaries led to considerable dilution of most fingerprint compounds. This in turn prompted a substantial reduction in contributions from all sources (the mainstream, tributary, and rain runoff) to 32, 12, and 8%, respectively. This reduction is also consistent with the qualitative analysis, which showed a significant decrease in the number of fingerprint compounds from all sources at this final site.

Despite the intensifying uncertainty associated with nontarget estimates as source concentrations decrease, yielding progressively fewer compounds to support estimates, nontarget methodologies prove capable of quantifying source contributions to systems that are relatively dilute, even when the majority of known contaminants fall below the detection limits. This observation underlines the potential applicability of nontarget methodologies to enhance our understanding of the contaminant distribution within complex environmental systems. It is important to emphasize that our adopted approach is based on the presumption that pollutant dispersion in the environment follows a dilution rule. Yet, the actual behavior of pollutants can be subject to a multitude of influences, including hydrodynamics, climatic conditions, topographical attributes, and other environmental parameters. These factors present a rich field for future research and need to be incorporated into our understanding to provide a more comprehensive and accurate depiction of pollutant behavior and distribution.

Environmental Implications. This study has far-reaching environmental implications, laying the groundwork for establishing fingerprints of various contaminant sources in real-world situations and subsequently employing them for both qualitative and quantitative analyses. Our research underscores the immense environmental insights that can be harvested from nontarget HRMS data. We initially identified novel contaminants in water samples and attempted to utilize these compounds for source tracking. However, we discovered that nontarget HRMS data deliver a more robust performance for both qualitative and quantitative analyses. By leveraging the wealth of this data, we bypassed the necessity for targeting individual contaminants. Instead, we applied an innovative method to quantitatively estimate the contributions of chemical sources to multifaceted mixed systems. Peering into the future, our methodology promises to be an invaluable asset in situations involving downstream contamination. It would be feasible, for example, to gather samples from all potential sources (such as wastewater from suspected factories), establish their unique fingerprints, and evaluate the contribution of each source to the contamination. By doing so, our approach presents itself as a comprehensive and pragmatic tool for environmental management and contaminant source tracking.

Moreover, the versatility of our approach extends beyond addressing contamination issues. It can serve as a valuable tool in ecological conservation efforts by tracking the source of nonpoint source pollutants, such as agricultural runoff, and aiding in the formulation of more effective pollution control strategies. Furthermore, our methodology can contribute to the development of better policies and regulations by providing more accurate data about pollution sources. It could also help identify emerging contaminants from different sources, allowing for early detection and mitigation. Importantly, while our technique has demonstrated potential, it should be acknowledged that it is still in its developmental phase. We intend this work to be the first in a series of studies aimed at refining our method and applying it to increasingly complex real-world scenarios. We envision that future research will explore the influence of various factors such as matrix effects, the persistence of nontarget compounds, and temporal variations in source composition on the accuracy of source apportionment. These endeavors will enhance our comprehension of intricate environmental systems, improve our ability to manage these systems effectively, and ultimately aid in the preservation of our environment.

■ ASSOCIATED CONTENT

Data Availability Statement

Data will be made available on request at https://figshare.com/articles/dataset/Raw_data_for_publication_zipRaw_dataRaw_data_for_publication/24260278.

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c06688.

Sampling site locations; examples of identification confidence levels; distribution and prevalence of identified CECs; information on isotope-labeled internal standards, detailed information regarding the 254 identifications, semiquantitative data of 46 compounds, and tentative compositions of the different fingerprints (PDF)

Detailed information regarding the 254 identifications (including level 1, 2, and 4), and the tentative compositions of the different fingerprints for C3 (n = 1689), T2 (n = 1317), and R1 (n = 15759) (XLSX)

AUTHOR INFORMATION

Corresponding Author

Rui Wang − *South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China; State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China;* orcid.org/0009-0002-4043-6979; Email: wangrui@scies.org

Authors

Di Xia − *South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China; State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China*

Lijun Liu − *South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China; State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China*

Bo Zhao − *South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China; State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China*

Danping Xie − *South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China; State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China*

Guining Lu − *School of Environment and Energy, South China University of Technology, Guangzhou 510006, China;* orcid.org/0000-0003-1974-1274

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.3c06688

Author Contributions

D.X.: study execution, data analysis, manuscript writing. L.L.: data analysis, compounds quantification. B.Z.: sampling sites selection, data analysis review. D.X.: data analysis review and manuscript editing. G.L.: data analysis review and manuscript editing. R.W.: study design, PyHRMS development, manuscript editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was funded wholly or in part by the Central Government Fund Supporting Nonprofit Scientific Institutes for Basic Research and Development (grant nos. PM-zx703-202112-332, PM-zx703-202104-112, PM-zx703-202011-321, PM-zx703-202204-112) and Guangdong Basic and Applied Basic Research Fund (no. 2021A1515110551).

REFERENCES

(1) Albergamo, V.; Schollée, J. E.; Schymanski, E. L.; Helmus, R.; Timmer, H.; Hollender, J.; de Voogt, P. Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. *Environ. Sci. Technol.* **2019**, *53* (13), 7584−7594.

(2) Barrett, H.; Du, X.; Houde, M.; Lair, S.; Verreault, J.; Peng, H. Suspect and Nontarget Screening Revealed Class-Specific Temporal Trends (2000−2017) of Poly- and Perfluoroalkyl Substances in St. Lawrence Beluga Whales. *Environ. Sci. Technol.* **2021**, *55* (3), 1659−1671.

(3) Bradley, P. M.; Journey, C. A.; Romanok, K. M.; Barber, L. B.; Buxton, H. T.; Foreman, W. T.; Furlong, E. T.; Glassmeyer, S. T.; Hladik, M. L.; Iwanowicz, L. R.; Jones, D. K.; Kolpin, D. W.; Kuivila, K. M.; Loftin, K. A.; Mills, M. A.; Meyer, M. T.; Orlando, J. L.; Reilly, T. J.; Smalling, K. L.; Villeneuve, D. L. Expanded Target-Chemical Analysis Reveals Extensive Mixed-Organic-Contaminant Exposure in U.S. Streams. *Environ. Sci. Technol.* **2017**, *51* (9), 4792−4802.

(4) Wang, S.; Basijokaite, R.; Murphy, B. L.; Kelleher, C. A.; Zeng, T. Combining Passive Sampling with Suspect and Nontarget Screening to Characterize Organic Micropollutants in Streams Draining Mixed-Use Watersheds. *Environ. Sci. Technol.* **2022**, *56* (23), 16726−16736.

(5) Been, F.; Kruve, A.; Vughs, D.; Meekel, N.; Reus, A.; Zwartsen, A.; Wessel, A.; Fischer, A.; ter Laak, T.; Brunner, A. M. Risk-Based Prioritization of Suspects Detected in Riverine Water Using Complementary Chromatographic Techniques. *Water Res.* **2021**, *204*, No. 117612.

(6) Alonso, M. B.; Maruya, K. A.; Dodder, N. G.; Lailson-Brito, J.; Azevedo, A.; Santos-Neto, E.; Torres, J. P. M.; Malm, O.; Hoh, E. Nontargeted Screening of Halogenated Organic Compounds in Bottlenose Dolphins (*Tursiops truncatus*) from Rio de Janeiro, Brazil. *Environ. Sci. Technol.* **2017**, *51* (3), 1176−1185.

(7) Pies, C.; Hoffmann, B.; Petrowsky, J.; Yang, Y.; Ternes, T. A.; Hofmann, T. Characterization and Source Identification of Polycyclic Aromatic Hydrocarbons (PAHs) in River Bank Soils. *Chemosphere* **2008**, *72* (10), 1594−1601.

(8) Höhener, P.; Guers, D.; Malleret, L.; Boukaroum, O.; Martin-Laurent, F.; Masbou, J.; Payraudeau, S.; Imfeld, G. Multi-Elemental Compound-Specific Isotope Analysis of Pesticides for Source Identification and Monitoring of Degradation in Soil: A Review. *Environ. Chem. Lett.* **2022**, *20* (6), 3927−3942.

(9) Liang, J.; Feng, C.; Zeng, G.; Gao, X.; Zhong, M.; Li, X.; Li, X.; He, X.; Fang, Y. Spatial Distribution and Source Identification of Heavy Metals in Surface Soils in a Typical Coal Mine City, Lianyuan, China. *Environ. Pollut.* **2017**, *225*, 681−690.

(10) Du, B.; Lofton, J. M.; Peter, K. T.; Gipe, A. D.; James, C. A.; McIntyre, J. K.; Scholz, N. L.; Baker, J. E.; Kolodziej, E. P. Development of Suspect and Non-Target Screening Methods for Detection of Organic Contaminants in Highway Runoff and Fish Tissue with High-Resolution Time-of-Flight Mass Spectrometry. *Environ. Sci.: Processes Impacts* **2017**, *19* (9), 1185−1196.

(11) Köppe, T.; Jewell, K. S.; Dietrich, C.; Wick, A.; Ternes, T. A. Application of a Non-Target Workflow for the Identification of

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Specific Contaminants Using the Example of the Nidda River Basin. *Water Res.* **2020**, *178*, No. 115703.
(12) Parry, E.; Young, T. M. Comparing Targeted and Non-Targeted High-Resolution Mass Spectrometric Approaches for Assessing Advanced Oxidation Reactor Performance. *Water Res.* **2016**, *104*, 72−81.
(13) Lara-Martín, P. A.; Chiaia-Hernández, A. C.; Biel-Maeso, M.; Baena-Nogueras, R. M.; Hollender, J. Tracing Urban Wastewater Contaminants into the Atlantic Ocean by Nontarget Screening. *Environ. Sci. Technol.* **2020**, *54* (7), 3996−4005.
(14) Chen, W. L.; Cheng, J. Y.; Lin, X. Q. Systematic Screening and Identification of the Chlorinated Transformation Products of Aromatic Pharmaceuticals and Personal Care Products Using High-Resolution Mass Spectrometry. *Sci. Total Environ.* **2018**, *637*−*638*, 253−263.
(15) Zahn, D.; Mucha, P.; Zilles, V.; Touffet, A.; Gallard, H.; Knepper, T. P.; Frömel, T. Identification of Potentially Mobile and Persistent Transformation Products of REACH-Registered Chemicals and Their Occurrence in Surface Waters. *Water Res.* **2019**, *150*, 86−96.
(16) Tian, Z.; Peter, K. T.; Gipe, A. D.; Zhao, H.; Hou, F.; Wark, D. A.; Khangaonkar, T.; Kolodziej, E. P.; James, C. A. Suspect and Nontarget Screening for Contaminants of Emerging Concern in an Urban Estuary. *Environ. Sci. Technol.* **2020**, *54* (2), 889−901.
(17) Renaguli, A.; Fernando, S.; Hopke, P. K.; Holsen, T. M.; Crimmins, B. S. Nontargeted Screening of Halogenated Organic Compounds in Fish Fillet Tissues from the Great Lakes. *Environ. Sci. Technol.* **2020**, *54* (23), 15035−15045.
(18) Goto, A.; Tue, N. M.; Isobe, T.; Takahashi, S.; Tanabe, S.; Kunisue, T. Nontarget and Target Screening of Organohalogen Compounds in Mussels and Sediment from Hiroshima Bay, Japan: Occurrence of Novel Bioaccumulative Substances. *Environ. Sci. Technol.* **2020**, *54* (9), 5480−5488.
(19) Gong, S.; Ren, K.; Ye, L.; Deng, Y.; Su, G. Suspect and Nontarget Screening of Known and Unknown Organophosphate Esters (OPEs) in Soil Samples. *J. Hazard. Mater.* **2022**, *436*, No. 129273.
(20) Brack, W.; Escher, B. I.; Müller, E.; Schmitt-Jansen, M.; Schulze, T.; Slobodnik, J.; Hollert, H. Towards a Holistic and Solution-Oriented Monitoring of Chemical Status of European Water Bodies: How to Support the EU Strategy for a Non-Toxic Environment? *Environ. Sci. Eur.* **2018**, *30*, No. 33.
(21) Hollender, J.; van Bavel, B.; Dulio, V.; Farmen, E.; Furtmann, K.; Koschorreck, J.; Kunkel, U.; Krauss, M.; Munthe, J.; Schlabach, M.; Slobodnik, J.; Stroomberg, G.; Ternes, T.; Thomaidis, N. S.; Togola, A.; Tornero, V. High Resolution Mass Spectrometry-Based Non-Target Screening Can Support Regulatory Environmental Monitoring and Chemicals Management. *Environ. Sci. Eur.* **2019**, *31*, No. 42.
(22) Lübeck, J. S.; Alexandrino, G. L.; Christensen, J. H. GC × GC−HRMS Nontarget Fingerprinting of Organic Micropollutants in Urban Freshwater Sediments. *Environ. Sci. Eur.* **2020**, *32*, No. 78.
(23) Xie, P.; Chen, S.; Liang, Y.; Wang, X.; Tian, R.; Upton, R. Chromatographic Fingerprint Analysis—a Rational Approach for Quality Assessment of Traditional Chinese Herbal Medicine. *J. Chromatogr. A* **2006**, *1112* (1−2), 171−180.
(24) Callao, M. P.; Ruisánchez, I. An Overview of Multivariate Qualitative Methods for Food Fraud Detection. *Food Control* **2018**, *86*, 283−293.
(25) Frank, R. A.; Roy, J. W.; Bickerton, G.; Rowland, S. J.; Headley, J. V.; Scarlett, A. G.; West, C. E.; Peru, K. M.; Parrott, J. L.; Conly, F. M.; Hewitt, L. M. Profiling Oil Sands Mixtures from Industrial Developments and Natural Groundwaters for Source Identification. *Environ. Sci. Technol.* **2014**, *48* (5), 2660−2670.
(26) Dávila-Santiago, E.; Shi, C.; Mahadwar, G.; Medeghini, B.; Insinga, L.; Hutchinson, R.; Good, S.; Jones, G. D. Machine Learning Applications for Chemical Fingerprinting and Environmental Source Tracking Using Non-Target Chemical Data. *Environ. Sci. Technol.* **2022**, *56* (7), 4080−4090.
(27) Peter, K. T.; Wu, C.; Tian, Z.; Kolodziej, E. P. Application of Nontarget High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. *Environ. Sci. Technol.* **2019**, *53* (21), 12257−12268.
(28) Peter, K. T.; Kolodziej, E. P.; Kucklick, J. R. Assessing Reliability of Non-Targeted High-Resolution Mass Spectrometry Fingerprints for Quantitative Source Apportionment in Complex Matrices. *Anal. Chem.* **2022**, *94* (6), 2723−2731.
(29) Xia, D.; Liu, H.; Lu, Y.; Liu, Y.; Liang, J.; Xie, D.; Lu, G.; Qiu, J.; Wang, R. Utility of a Non-Target Screening Method to Explore the Chlorination of Similar Sulfonamide Antibiotics: Pathways and N-Cl Intermediates. *Sci. Total Environ.* **2023**, *858*, No. 160042.
(30) Wang, R.; Yan, Y.; Liu, H.; Li, Y.; Jin, M.; Li, Y.; Tao, R.; Chen, Q.; Wang, X.; Zhao, B.; Xie, D. Integrating Data Dependent and Data Independent Non-Target Screening Methods for Monitoring Emerging Contaminants in the Pearl River of Guangdong Province, China. *Sci. Total Environ.* **2023**, *891*, No. 164445.
(31) Yang, X.; Wang, R.; He, Z.; Dai, X.; Jiang, X.; Liu, H.; Li, Y. Abiotic Transformation of Synthetic Progestins in Representative Soil Mineral Suspensions. *J. Environ. Sci.* **2023**, *127*, 375−388.
(32) Liang, J.; Wang, R.; Liu, H.; Xie, D.; Tao, X.; Zhou, J.; Yin, H.; Dang, Z.; Lu, G. Unintentional Formation of Mixed Chloro-Bromo Diphenyl Ethers (PBCDEs), Dibenzo-p-Dioxins and Dibenzofurans (PBCDD/Fs) from Pyrolysis of Polybrominated Diphenyl Ethers (PBDEs). *Chemosphere* **2022**, *308*, No. 136246.
(33) Jiang, X.; Xue, Z.; Chen, W.; Xu, M.; Liu, H.; Liang, J.; Zhang, L.; Sun, Y.; Liu, C.; Yang, X. Biotransformation Kinetics and Pathways of Typical Synthetic Progestins in Soil Microcosms. *J. Hazard. Mater.* **2023**, *446*, No. 130684.
(34) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.
(35) Aalizadeh, R.; Alygizakis, N.; Schymanski, E.; Slobodnik, J. *Norman Network*, S0 | SUSDAT | 619 Merged NORMAN Suspect List: SusDat, 2018.
(36) MassBank | MassBank of North America Mass Spectral DataBase, 2022 https://mona.fiehnlab.ucdavis.edu. Accessed July 15, 2022.
(37) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097−2098.
(38) Liu, H.; Wang, R.; Zhao, B.; Xie, D. Assessment for the Data Processing Performance of Non-Target Screening Analysis Based on High-Resolution Mass Spectrometry. *Sci. Total Environ.* **2024**, *908*, No. 167967.
(39) Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Software* **2021**, *6* (60), No. 3021.
(40) Gu, Z.; Eils, R.; Schlesner, M. Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics* **2016**, *32* (18), 2847−2849.
(41) Ort, C.; Lawrence, M. G.; Reungoat, J.; Eaglesham, G.; Carter, S.; Keller, J. Determining the Fraction of Pharmaceutical Residues in Wastewater Originating from a Hospital. *Water Res.* **2010**, *44* (2), 605−615.
(42) García, M. G.; Sánchez, J. I. L.; Bravo, K. A. S.; Cabal, M. D. C.; Pérez-Santín, E. Review: Presence, Distribution and Current Pesticides Used in Spanish Agricultural Practices. *Sci. Total Environ.* **2022**, *845*, No. 157291.
(43) Beck, J.; Brüggemann, M.; Van Pinxteren, D.; Herrmann, H. Nontarget Approach to Identify Complexing Agents in Atmospheric

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Aerosol and Rainwater Samples. *Anal. Chem.* **2022**, *94* (25), 8966−8974.
(44) Jurado, A.; Pujades, E.; Walther, M.; Diaz-Cruz, M. S. Occurrence, Fate, and Risk of the Organic Pollutants of the Surface Water Watch List in European Groundwaters: A Review. *Environ. Chem. Lett.* **2022**, *20* (5), 3313−3333.
(45) Wang, T.; Zhong, M.; Lu, M.; Xu, D.; Xue, Y.; Huang, J.; Blaney, L.; Yu, G. Occurrence, Spatiotemporal Distribution, and Risk Assessment of Current-Use Pesticides in Surface Water: A Case Study near Taihu Lake, China. *Sci. Total Environ.* **2021**, *782*, No. 146826.
(46) Hou, L.; Jin, X.; Liu, N.; Luo, Y.; Yan, Z.; Chen, M.; Liu, Y.; Xie, H.; Giesy, J. P.; Wu, F.; Xu, J. Triadimefon in Aquatic Environments: Occurrence, Fate, Toxicity, and Ecological Risk. *Environ. Sci. Eur.* **2022**, *34* (1), No. 12.
(47) Akhtar, N.; Ishak, M. I. S.; Bhawani, S. A.; Umar, K. Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review. *Water* **2021**, *13* (19), No. 2660.
(48) Rose, L. D.; Akob, D. M.; Tuberty, S. R.; Corsi, S. R.; DeCicco, L. A.; Colby, J. D.; Martin, D. J. Use of High-Throughput Screening Results to Prioritize Chemicals for Potential Adverse Biological Effects within a West Virginia Watershed. *Sci. Total Environ.* **2019**, *677*, 362−372.
(49) Masoner, J. R.; Kolpin, D. W.; Cozzarelli, I. M.; Barber, L. B.; Burden, D. S.; Foreman, W. T.; Forshay, K. J.; Furlong, E. T.; Groves, J. F.; Hladik, M. L.; Hopton, M. E.; Jaeschke, J. B.; Keefe, S. H.; Krabbenhoft, D. P.; Lowrance, R.; Romanok, K. M.; Rus, D. L.; Selbig, W. R.; Williams, B. H.; Bradley, P. M. Urban Stormwater: An Overlooked Pathway of Extensive Mixed Contaminants to Surface and Groundwaters in the United States. *Environ. Sci. Technol.* **2019**, *53* (17), 10070−10081.
(50) Podder, A.; Sadmani, A. H. M. A.; Reinhart, D.; Chang, N. B.; Goel, R. Per and Poly-Fluoroalkyl Substances (PFAS) as a Contaminant of Emerging Concern in Surface Water: A Transboundary Review of Their Occurrences and Toxicity Effects. *J. Hazard. Mater.* **2021**, *419*, No. 126361.
(51) Llamas-Dios, M. I.; Vadillo, I.; Jiménez-Gavilán, P.; Candela, L.; Corada-Fernández, C. Assessment of a Wide Array of Contaminants of Emerging Concern in a Mediterranean Water Basin (Guadalhorce River, Spain): Motivations for an Improvement of Water Management and Pollutants Surveillance. *Sci. Total Environ.* **2021**, *788*, No. 147822.
(52) Carpenter, C. M. G.; Wong, L. Y. J.; Johnson, C. A.; Helbling, D. E. Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. *Environ. Sci. Technol.* **2019**, *53* (1), 77−87.
(53) Zhang, X.; Lohmann, R.; Dassuncao, C.; Hu, X. C.; Weber, A. K.; Vecitis, C. D.; Sunderland, E. M. Source Attribution of Poly- and Perfluoroalkyl Substances (PFASs) in Surface Waters from Rhode Island and the New York Metropolitan Area. *Environ. Sci. Technol. Lett.* **2016**, *3* (9), 316−321.
(54) Kibbey, T. C. G.; Jabrzemski, R.; O'Carroll, D. M. Supervised Machine Learning for Source Allocation of Per- and Polyfluoroalkyl Substances (PFAS) in Environmental Samples. *Chemosphere* **2020**, *252*, No. 126593.
(55) Peter, K. T.; Tian, Z.; Wu, C.; Lin, P.; White, S.; Du, B.; McIntyre, J. K.; Scholz, N. L.; Kolodziej, E. P. Using High-Resolution Mass Spectrometry to Identify Organic Contaminants Linked to Urban Stormwater Mortality Syndrome in Coho Salmon. *Environ. Sci. Technol.* **2018**, *52* (18), 10317−10327.
(56) Nürenberg, G.; Kunkel, U.; Wick, A.; Falås, P.; Joss, A.; Ternes, T. A. Nontarget Analysis: A New Tool for the Evaluation of Wastewater Processes. *Water Res.* **2019**, *163*, No. 114842.
(57) Stahnke, H.; Kittlaus, S.; Kempe, G.; Alder, L. Reduction of Matrix Effects in Liquid Chromatography−Electrospray Ionization−Mass Spectrometry by Dilution of the Sample Extracts: How Much Dilution Is Needed? *Anal. Chem.* **2012**, *84* (3), 1474−1482.
(58) Boysen, A. K.; Heal, K. R.; Carlson, L. T.; Ingalls, A. E. Best-Matched Internal Standard Normalization in Liquid Chromatography−Mass Spectrometry Metabolomics Applied to Environmental Samples. *Anal. Chem.* **2018**, *90* (2), 1363−1369.
(59) Ng, B.; Quinete, N.; Gardinali, P. R. Assessing Accuracy, Precision and Selectivity Using Quality Controls for Non-Targeted Analysis. *Sci. Total Environ.* **2020**, *713*, No. 136568.

https://doi.org/10.1021/acs.est.3c06688
Environ. Sci. Technol. XXXX, XXX, XXX−XXX