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

Kelly L. Pereira,* Martyn W. Ward, John L. Wilkinson, Jonathan Brett Sallach, Daniel J. Bryant, William J. Dixon, Jacqueline F. Hamilton, and Alastair C. Lewis

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The life-critical matrices of air and water are among the most complex chemical mixtures that are ever encountered. Ultrahigh-resolution mass spectrometers, such as the Orbitrap, provide unprecedented analytical capabilities to probe the molecular composition of such matrices, but the extraction of non-targeted chemical information is impractical to perform via manual data processing. Automated non-targeted tools rapidly extract the chemical information of all detected compounds within a sample dataset. However, these methods have not been exploited in the environmental sciences. Here, we provide an automated and (for the first time) rigorously tested methodology for the non-targeted compositional analysis of environmental matrices using coupled liquid chromatography—mass spectrometric data. First, the robustness and reproducibility was tested using authentic standards, evaluating performance as a function of concentration, ionization potential, and sample complexity. The method was then used for the compositional analysis of particulate matter and surface waters collected from worldwide locations. The method detected >9600 compounds in the individual environmental samples, arising from critical pollutant sources, including carcinogenic industrial chemicals, pesticides, and pharmaceuticals among others. This methodology offers considerable advances in the environmental sciences, providing a more complete assessment of sample compositions while significantly increasing throughput.

KEYWORDS: ultrahigh-resolution mass spectrometry, non-targeted analysis, Compound Discoverer, liquid chromatography—mass spectrometry

INTRODUCTION

Environmental pollution accounts for ~9 million premature deaths per annum.[1] Of these deaths, 4.2 million are attributed to ambient particulate matter (PM) and a further 1.8 million to unsafe water sources and sanitation.[2] Air and water matrices typically contain 10^3-10^5 pollutants, with a diverse range of chemical functionalities and concentrations.[3,4] The sheer number of pollutants present in the environment, their varied sources, chemical functionalities, and concentrations make the compositional analysis of these species a formidable analytical challenge. Regulatory and enforcement bodies assess air and water pollution via the measurement of prescribed lists of targeted compounds using predefined analytical methods.[5,6] While the number of regulated pollutants has increased in recent years, these targeted compounds represent a tiny proportion of the thousands of pollutants actually present. Consequently, most pollutants in air and water go undetected, including potentially hazardous site-specific and emerging contaminants.[7]

Fourier transform mass spectrometers (FTMS) offer unprecedented capabilities to probe the molecular composition of highly complex matrices, with resolving powers greater than ~10^5 full width at half-maximum at a mass-to-charge ratio (m/z) of 200, with mass accuracies <2 parts per million. There are only two FTMS: the Orbitrap (ThermoFisher Scientific) and the Fourier transform ion-cyclotron resonance (Bruker Daltonics). The molecular identification of unknowns in
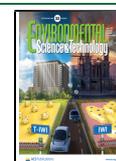
complex matrices requires a mass analyzer with high mass accuracy to reduce the ambiguity in molecular formula assignments and high mass resolution to minimize the overlap of isobaric species. Chromatographic separation provides additional chemical specificity, allowing structural isomers to be distinguished. These techniques however generate significant volumes of highly complex data, resulting in users targeting limited lists of compounds to reduce data complexity and increase throughput. These targeted approaches, analogous to regulatory methods, provide severely limited compositional information.

Automated non-targeted screening tools can overcome these challenges, rapidly extracting the chemical information of all detected compounds within a sample dataset, highlighting background artifacts, providing molecular formula assignments (among other information) and probable structure through mass spectral library screening. This compositional information is incredibly beneficial and essential in several scientific disciplines.[8−12] While automated non-targeted screening can reduce data analysis time from months to hours, these methods have not been exploited in the environmental sciences. Non-targeted tools have been designed mainly for the analysis of metabolites and proteins ([13−16] and references therein). These tools lack the chemical metrics frequently used in the environmental sciences to aid in the identification of pollutant sources (e.g., elemental ratios,[17] average carbon oxidation state,[18] aromaticity index,[19] and various compositional groupings[20−22]) and may in part account for the slow uptake within the environmental domain. Further, to our knowledge, no studies have investigated the performance and reproducibility of these methods for the compositional analysis of trace-level compounds in environmental matrices.

Here, we present an automated methodology for the non-targeted compositional analysis of environmental matrices using coupled ultrahigh-performance liquid chromatography−Orbitrap MS. The method consists of a bespoke workflow developed in the instrument manufacturers' commercial software, providing seamless integration with the instrumentation and automated screening of the largest tandem mass spectral database (mzCloud, www.mzcloud.org; not possible with other platforms) and a custom-built data processing program. The data program further advances the workflow capabilities, including a more rigorous approach for the removal of artifacts from the sample data and the automated calculation of numerous environmental chemical metrics and groupings to aid in compositional interpretation and allow for the rapid comparison of sample compositions. The method is applicable to any ThermoFisher Scientific FTMS (i.e., FTMS market leader and includes all Orbitrap designs). First, we test the ability of the method to detect, identify, and integrate authentic standards frequently observed in environmental matrices, evaluating performance as a function of analyte concentration, ionization potential, and sample complexity. We then evaluate the performance for the analyses of PM and surface waters collected from worldwide locations.

MATERIALS AND METHODS

Standards. Sixty authentic standards were used to test the detection, identification, and integration capabilities of the non-targeted data processing method. The compound names, manufacturer, and purity of the standards can be found in Table S1. The standards were prepared at 1 ppm and in mixtures at concentrations ranging from 5 ppm to 0.5 ppb.

Two compounds, furan-2,5-dione and 3-methylfuran-2,5-dione, were excluded from the standard mixture to prevent the formation of their acid counterparts, which were also included. Standards were prepared in 50:50 methanol/water (optima, LC−MS grade, ThermoFisher Scientific) for analysis. Calibrations were performed for any of the standards identified in the environmental samples and consisted of a minimum of five concentrations, with three replicate measurements per concentration.

Ambient Particulate Matter. Samples were collected at the Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences in Beijing, China (Lat. 39°58′28″ N, Long. 116°22′15″ E) onto quartz fiber filters at a flow rate of 1.33 m^3/min using a HiVol sampler with a PM$_{2.5}$ inlet (model 3000, Ecotech). The sampler was positioned on the roof of the IAP building ∼8 m above ground level. Filters were pre-conditioned in a furnace at 500 °C for 5 h to remove any volatiles before use. The sampling dates and times are shown in Table S2. After sample collection, each filter was wrapped in foil to minimize potential photolysis degradation and stored in a freezer at −20 °C. Filters were shipped in dry ice to the University of York for analysis. Samples were prepared using the methodology detailed in Bryant et al. (2019).[23] Briefly, 1/8th of each filter was extracted into 4 mL of water, left for 2 h at room temperature, and sonicated for 30 min. The aqueous extract was then filtered through a 0.22 μm filter membrane, evaporated to dryness (model V10, Biotage, Sweden), and reconstituted in 1 mL of 50:50 methanol/water. A procedural blank was also prepared, consisting of a blank pre-conditioned filter subjected to the same sample extraction and preparation procedure.

Surface Waters. Samples were collected from China, India, and Sri Lanka. Full details of the sample collection locations are shown in Table S3. Samples were collected using the protocol described in Wilkinson et al. (2019).[24] Briefly, 10 mL aliquots of water were collected via grab sampling, filtered through a 0.45 μm glass microfiber filter membrane, and shipped in dry ice to the University of York where they were stored at −80 °C. Prior to analysis, 800 μL of each sample was evaporated to dryness and reconstituted in 80 μL of 50:50 methanol/water. A procedural blank was also prepared consisting of 800 μL of high-purity water, subjected to the same sample preparation procedure.

Instrument and Data Analysis. Full method details can be found in the Supporting Information; a brief summary is given here. The standards and samples were analyzed using ultrahigh-performance liquid chromatography coupled to an ultrahigh-resolution mass spectrometer (Dionex 3000-QExactive Orbitrap, UPLC-MS, ThermoFisher Scientific). Data were analyzed using the non-targeted method, consisting of a bespoke workflow developed in the framework Compound Discoverer (version 2.1, ThermoFisher Scientific) and a custom-built data processing program developed in Python (version 3.7). The workflow is shown in Figure S1. The workflow was designed to, (i) align the chromatographic retention times of input data files, (ii) extract all chromatographic peaks which met set criteria (see below), (iii) assign the molecular formula for each compound and, (iv) screen the MS^2 data against an in-house built and commercial library (mzCloud) for possible compound identification. The in-house MS^2 library was developed in the software package mzVault (version 2.0, ThermoFisher Scientific, supplied with Compound Discoverer) using the spectra obtained from the

analysis of the 60 individually prepared standards. Chromatographic peaks were detected if the molecular species had a signal-to-noise ratio >3, a minimum peak intensity of 3×10^4 and was detected in a minimum of three consecutive scans. Molecular formula assignments were allowed unlimited C, H, O atoms and up to 5 N atoms, 2 S atoms and 3 Cl atoms (surface water analysis only). In positive ionization mode, 2 Na atoms and 1 K atom were also allowed. Molecular formulae were only assigned if the isotopic intensity tolerance was within ±30% of the theoretical isotopic abundance and the mass tolerance was <3 ppm. The software also screens the sample data for the detection of common electrospray ionization (ESI) artifacts. The list of ESI artifacts is shown in Table S4. Where multiple adducts are detected, the software will group these species and report the data for only one adduct, typically $[M + H]^+$ or $[M - H]^-$ (i.e., user-specified preferred adduct). All other artifacts are removed from the sample data. Specific workflows were developed for the analysis of positive and negative ionization mode data. The data program was developed to perform additional screening functions and calculations, which could not be performed in Compound Discoverer. The program was designed to (i) tabulate the workflow output into a user-friendly format, (ii) remove system (i.e., solvent blanks) and sample preparation (i.e., method procedural blanks) artifacts using a more rigorous approach (designed for complex matrices, see the Supporting Information, "Removal of Artifacts"), (iii) remove components with unassigned or erroneous molecular formulae, (iv) perform chemical metric calculations to aid in the identification of pollutant sources, and (v) output the data using various chemical groupings to allow for the rapid comparison of sample compositions. All manual data processing was performed in the software Freestyle (version 1.1, Thermo-Fisher Scientific). The workflows, in-house MS^2 library, and data program can be downloaded from a public depository (doi:10.5281/zenodo.4701800).

RESULTS

Initial Software Testing. Sixty authentic standards were initially used to test the detection, identification, and integration capabilities of the non-targeted method. The standards contained a diverse range of chemical functionalities (e.g., carboxylic acids, carbonyls, alcohols, aromatics, nitrophenols) representing the types of compounds often observed in environmental samples. The molecular weight (MW) of the standards ranged from 96 to 232, with an average oxygen-to-carbon (O/C) ratio of 0.47 and a carbon number range of C_3 to C_{15}. First, the ability of the method to detect and identify sample components was tested using the individually prepared standards at a concentration of 1 ppm. The sample complexity was then increased, testing the method's ability to detect, identify, and integrate the standards in a mixture prepared at concentrations between 5 ppm to 0.5 ppb. The standards were grouped into the types of molecular species detected (e.g., deprotonated, protonated, sodiated), investigating whether the performance was affected by negative ionization mode or adduct formation in positive ionization mode.

The 60 standards were initially characterized via manual data processing, recording if the standard was detected, the ionization mode and type of molecular species, the chromatographic retention time, and the MS^2 fragmentation spectrum, which was recorded in the in-house library. This data was then used to evaluate the performance of the non-targeted method.

In total, 45 standards were detected in negative ionization mode as deprotonated molecular species $[M - H]^-$. In positive ionization mode, 28 standards were detected as protonated molecular species $[M + H]^+$ and a further 26 standards were detected as sodiated molecular species $[M + Na]^+$. Potassiated molecular species $[M + K]^+$ were observed for some standards but represented less than ~1% of the total precursor signal intensity and were subsequently excluded from further analysis. The detected molecular species and their retention times (determined via manual analysis) are shown in Table S5.

The performance of the non-targeted method to detect the chromatographic peaks and assign the molecular formulae and compound names (i.e., molecular identities, determined from MS^2 library screening) of the standards is shown in Figure S2. The non-targeted method successfully detected the chromatographic peaks and correctly identified the molecular formulae for all 45 $[M - H]^-$ and 28 $[M + H]^+$ standards (Figure S2A,B). The ability of the method to detect and identify $[M + Na]^+$ standards is shown in Figure S2C. The method detected the chromatographic peaks and provided the molecular formulae for 15 out of 26 $[M + Na]^+$ standards (i.e., 58%). The software appears to initially search for $[M + H]^+$ species in positive ionization mode. If detected, the software then searches for $[M + Na]^+$ (see the Supporting Information, "Sodium Adduct Detection" for further information). Consequently, any compounds that are exclusively observed as $[M + Na]^+$ in positive ionization mode are not detected by the software. This parameterization cannot be changed by the user (i.e., a part of the underlying software algorithms), demonstrating the importance of testing automated non-targeted methods to provide a fundamental understanding of their potential limitations prior to use.

The number of standards identified by the non-targeted method using the in-house and commercial library is shown in Figure S2. The MS^2 spectrum of some standards was not obtained during analysis, preventing molecular identification using library screening (shown as "omitted" data in Figure S2). Excluding those standards where no MS^2 spectra were obtained, the method provided the molecular identity for 84% of the $[M - H]^-$ (34 out of a possible 44) and 95% of the $[M + H]^+$ (21 out of 22) standards. The software was unable to provide the molecular identities of the standards exclusively observed as $[M + Na]^+$ in positive ionization mode as the chromatographic peaks were not detected. Excluding these standards, the non-targeted method provided the molecular identity for 73% of the $[M + Na]^+$ standards (11 out of a possible 15). Overall, the in-house library provided the identities for 85% (69 out of 89) of the total number of molecular species (i.e., deprotonated, protonated, and sodiated standards). The commercial library contained spectra for 24 of the 60 standards in its database and subsequently provided the identity of fewer standards in comparison to the in-house library, identifying only 21% (18 out of 89) of the total number of molecular species.

Increasing Sample Complexity. The 60 standards (excluding two compounds, see Materials and Methods) were combined into a single mixture, prepared at various concentrations, and used as a proxy to test the performance of the method. The chromatographic peak areas of the individually prepared standards (at the same concentration) were integrated and used as a metric to describe the ionization efficiency of each molecular species, allowing the standards to
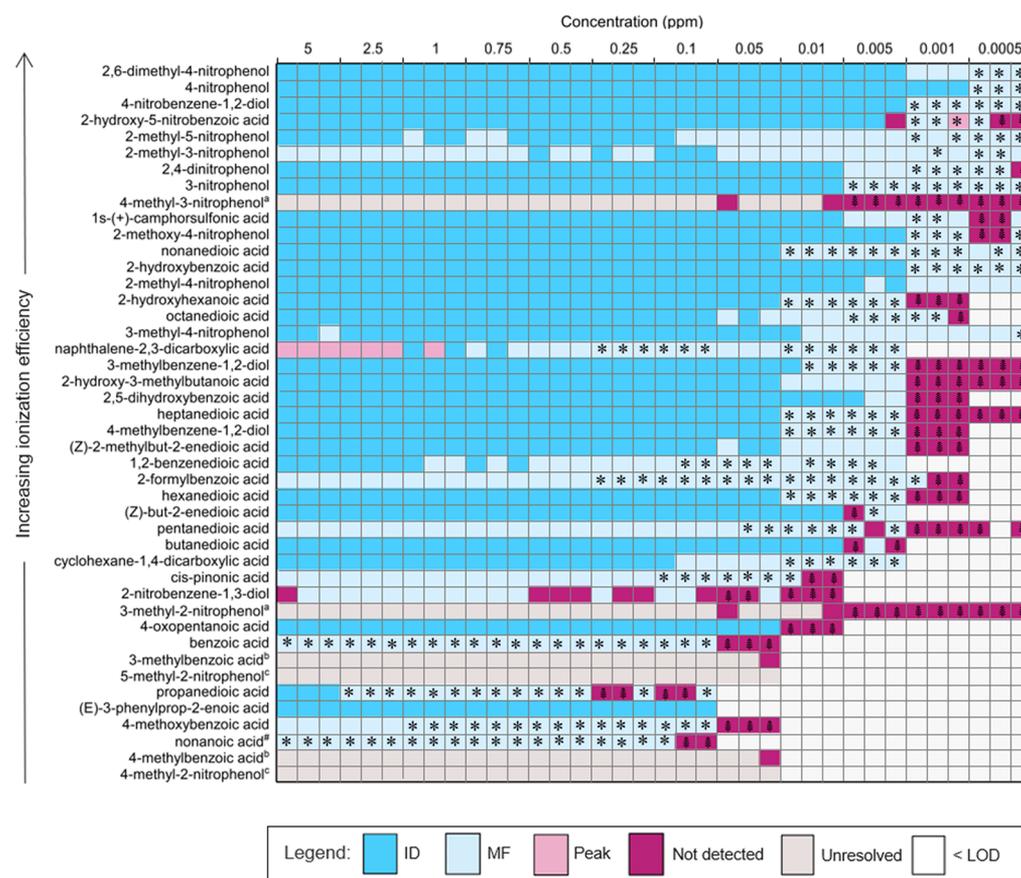
Figure 1. Performance of the non-targeted method to detect and identify $[M − H]^-$ species in the standard mix prepared at various concentrations. The plot displays whether the compound names (ID), molecular formulae (MF), and chromatographic peak (Peak) were identified. Each box represents one measurement, with three replicate measurements performed for each concentration. The asterisks indicate that no MS^2 data was recorded during analysis preventing molecular identification. The palm branches indicate that the chromatographic peak cannot be detected due to the use of unit mass resolution. Isomeric species that could not be resolved via manual or automated data processing are shown in gray. Letters correspond to the groups of isomeric species which could not be resolved; [a] = 3-methyl-2-nitrophenol and 4-methyl-3-nitrophenol. [b] = 3-methylbenzoic acid and 4-methylbenzoic acid, [c] = 5-methyl-2-nitrophenol and 4-methyl-2-nitrophenol. The hash symbol indicates that the in-house library contains no MS^2 spectra for this standard.

be ordered by increasing ionization efficiency. The ability of the method to detect and identify each compound in the standard mixtures, in negative and positive ionization mode, is shown in Figure 1 and Figure S3, respectively. Three replicate sample injections were performed for each standard mixture to investigate the reproducibility of the non-targeted method to report the same result. Several isomeric compounds could not be resolved in the standard mixtures via manual or automated data processing due to co-elution and are shown in Figure 1 and Figure S3 as "unresolved".

From Figure 1 and Figure S3A, it can be observed that the non-targeted method struggled to detect some standards at the lowest observable concentration and was particularly evident with decreasing ionization efficiency. There are four main parameters in the non-targeted method that control chromatographic peak detection, particularly at low concentrations. The majority of these parameters can be found in the "detect unknown compounds" node (Figure S1). However, the workflow is based on a hierarchical structure. Therefore, any parameters prior to and including the detect unknown compounds node can affect chromatographic peak detection. The four main parameters include the signal-to-noise (S/N) threshold (select spectra and detect unknown compounds node), minimum peak intensity, isotopic intensity tolerance,

and the minimum scans per peak. Each parameter was individually tested, increasing or decreasing the value to remove any restrictions. Excluding the S/N threshold in the select spectra node, which incorrectly determined some standards to be <LOD (see the Supporting Information, "Software Notes"), all other parameters did not improve the chromatographic peak detection capabilities.

The non-detection of the low concentration species can instead be explained by considering the mass resolution used for chromatographic peak detection. The software uses unit mass resolution (i.e., an integer value) for chromatographic peak detection. This is incredibly beneficial for reducing software processing time. However, this does not utilize the accurate mass capability of the instrumentation, capable of achieving >6 decimal places. The use of unit mass resolution (particularly for the analysis of complex matrices) will not be able to fully resolve chromatographic peaks from other components in the sample, resulting in lower S/N ratios, increasing the number of chromatographic peaks, which are determined to be <LOD. For example, deprotonated 3-methylbenzene-1,2-diol in the 0.001 ppm standard mixture had an S/N ratio of 1.89 using a unit mass range of m/z 123 to 124 (determined via manual analysis). Using an accurate mass range of m/z 123.0450 to 123.0460, deprotonated 3-
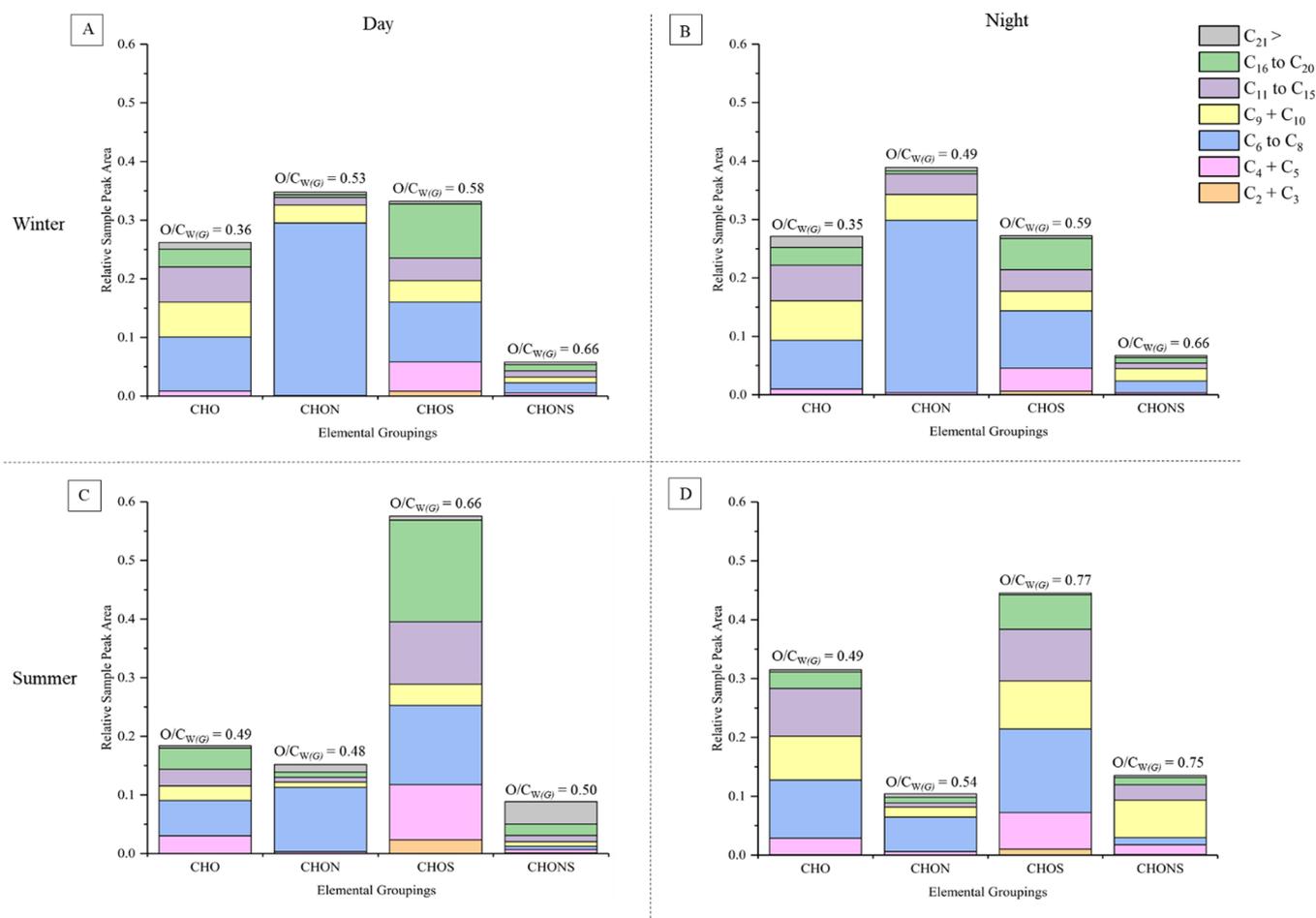
Figure 2. $PM_{2.5}$ samples collected in Beijing during the winter season in the day (A) and night (B) and summer season in the day (C) and night (D), displaying all detected compounds grouped by their elemental composition and number of carbon atoms in each molecular formula, as a function of their relative sample peak area. Each plot shows the average composition of two aerosol samples collected during the same season and time of day using the data acquired from negative ionization mode. $O/C_{W(G)}$ displays the weighted oxygen-to-carbon ratio, calculated by dividing the peak area of each elemental grouping by the total sample peak area.

methylbenzene-1,2-diol had an S/N ratio of 20.8 (a factor of 11 increase), accounting for the difficulties in the detection of low concentration species. Analogous to sodium adduct detection, the mass resolution used for chromatographic peak detection cannot be controlled by the user (a part of the underlying software algorithms).

The number of chromatographic peaks, which could not be detected using unit mass resolution, is shown in Figure 1 and Figure S3 (demonstrated using manual analysis). Unit mass resolution accounted for the non-detection of 86 and 76% of the $[M − H]^-$ and $[M + H]^+$ standards, respectively. It is worth noting that protonated 4-methoxybenzoic acid co-eluted with an isomeric fragment of (4-formyl-2-methoxyphenol)-acetate, accounting for the difficulty in the detection of this species. Excluding protonated 4-methoxybenzoic acid, unit mass resolution accounted for the non-detection of all $[M + H]^+$ standards. Conversely, only 26% of the non-detected $[M + Na]^+$ standards could be attributed to the use of unit mass resolution for chromatographic peak detection. The non-detected $[M + Na]^+$ standards (as previously discussed) is primarily due to the inability of the software to assign sodium adduct formation if the protonated molecular species is not detected, accounting for the non-detection of levoglucosan, 2-hydroxyhexanoic acid, 2-hydroxy-3-methylbutanoic acid, and

2,3-diacetyloxypropyl acetate, which were exclusively observed as $[M + Na]^+$ in positive ionization mode, as shown in Figure S3B.

A summary of the method's performance to detect and identify each compound in the standard mixtures is shown in Figure S4. Overall, the non-targeted method detected the chromatographic peaks and identified the molecular formulae for 91.6 ± 1.0% (mean ± variation from the mean) of the total number of $[M − H]^-$ and $[M + H]^+$ standards and provided the identity for 70 and 59%, respectively. For $[M + Na]^+$ standards, the method detected the chromatographic peaks and provided the molecular formulae for 57% and correctly identified 28%. The method consistently reported the same result for each standard in the replicate sample injection measurements and data analyses, with 92% of all molecular species displaying no variation (Figure 1 and Figure S3). The largest variation in the detection and identification of the standards was observed at the lowest concentrations, likely the result of low-intensity MS^2 spectra and/or compounds close to the LOD (i.e., instrument variation close to the software "cut-off" values). The integration capabilities of the non-targeted method are shown in Figure S5. The chromatographic peaks of the standards were integrated via the non-targeted method and manual data processing, allowing calibration graphs to be

plotted and the integration capabilities of the two methods to be compared. Both methods displayed good agreement with an R^2 of 0.9993 and a slope of 1.14 ± 0.007. The primary limitation observed with increasing sample complexity was poor chromatographic separation of isomeric species. Where two unresolved isomeric compounds were present, the non-targeted method typically reported the detection of only one of the standards, usually the most abundant. This limitation however cannot be attributed to the non-targeted method; those compounds could not be resolved using either manual or automated data processing and is ultimately dependent upon the analytical method (i.e., a balance between throughput and chromatographic resolution).

Analysis of Ambient Particulate Matter. Here, we use the non-targeted method to investigate the chemical composition of eight ambient PM samples collected in Beijing, China, evaluating the method's performance through the comparison of detected pollutants, their sources, and abundance (including quantitative measurements) with literature observations and modeled data. The method incorporates two screening approaches: (i) non-targeted, where the chemical information of all detected compounds in each sample is reported, and (ii) targeted, which uses the in-house library to screen the samples for the identification of the 60 authentic standards initially used to test the method. Manual data analysis was also used to test whether the non-targeted method had correctly reported the identification of the targeted standards in the PM samples. The molecular identification of these compounds was confirmed using the retention times and fragmentation spectra of the authentic standards. Calibrations were also performed, providing quantitative measurements. The PM samples consisted of four samples collected during the summer season and a further four collected during the winter season. Each respective season included two samples collected during the daytime and two overnight. Full details can be found in Table S2.

The non-targeted method detected between 4402 to 9655 compounds in the individual PM samples (sum of negative and positive ionization mode, see Table S6). The in-house library identified a total of 185 compounds, which included multiple identifications of the standards in each PM sample. Of the targeted compounds, 147 were quantified. The concentrations of the other 38 compounds could not be determined as the chromatographic peak areas were outside (i.e., either below or above) the measured linear calibration ranges preventing quantification. The ambient concentrations of the targeted compounds in the PM samples are shown in Tables S7 and S8. The sheer number of compounds, which can be detected using the non-targeted vs targeted approach, is shown in Figure S6, as an example for one PM sample. Of the 60 standards used for targeted identification, 20 were identified in this sample. In contrast, the non-targeted approach detected 5089 unique compounds (i.e., chemically and/or structurally different). Using the targeted approach, <1.1% of the organic PM mass (by weight) was quantified in all the samples, demonstrating the importance of using non-targeted methods in the environmental sciences to provide a more complete assessment of sample compositions.

To evaluate the insights, which can be obtained using non-targeted data, we compare the chemical composition of the ambient PM samples collected during summer and winter, day and night. The composition of each PM sample was separated into several groupings using the automated data processing

program, see the Supporting Information, "Data Processing Program" for further information. Briefly, all detected compounds in each PM sample were grouped by the number of carbon atoms and elemental composition in their molecular formulae, relating to potential pollutant sources. The peak areas of each compound were normalized to the total peak area in each sample, allowing the relative abundance of the chemical groupings between samples to be compared, as shown in Figure 2.

Figure 2 shows that the abundance of the C_6 to C_8 CHON grouping is ~3 times more significant in winter vs summer. Using the processed program data, the chemical information of the compounds, which contribute to this grouping in both PM samples, can be rapidly observed. In the winter daytime sample (sample 96, see Table S2), 220 compounds were detected in the C_6 to C_8 CHON grouping. 98% of these compounds had DBE/C values >0.5, suggesting that the majority are aromatic and polycyclic aromatic compounds, also supported by the commercial and in-house library matches. The compounds were relatively oxidized, with an average O/C ratio of 0.52 and MW of 177. In contrast, fewer compounds were observed in the C_6 to C_8 CHON grouping in the daytime summer sample (sample 261). In this sample, 77 compounds were detected, 88% which had DBE/C values >0.5. The average O/C ratio was 0.59 and MW of 184, suggesting that the winter and summer PM sample compositions are relatively similar, with the summer sample containing more oxidized and non-aromatic compounds.

Using chemical metric plots, the compositional differences between these chemical groupings can be further explored. Figure S7 shows the composition and relative abundance of each C_6 to C_8 CHON compound in the summer and winter samples in a DBE/C vs molecular weight space. While the winter sample contained a greater number of C_6 to C_8 CHON species, Figure S7 shows that the composition is dominated by ~3 compounds, which have considerably higher abundances than in the summer sample, accounting for the differences in Figure 2. The relative abundance of several identified compounds (determined via targeted screening), including 4-nitrophenol, 3-methyl-4-nitrophenol, and 2,6-dimethyl-4-nitrophenol were comparable between winter and summer, with ratios of 1.3, 6.0, and 3.0 for winter/summer, respectively. The most abundant compound in winter was identified as 4-nitrobenzene-1,2-diol (i.e., 4-nitrocatechol). The other two abundant compounds are suggested to be methyl nitro-catechols (t_R 7.15 and 8.82), displaying the characteristic neutral losses of NO, HNO_2, or NO_2 and the combined loss of NO and CO.[25,26] 4-Nitrocatechol and methyl nitrocatechols are well-known abundant oxidation products of biomass burning, formed via OH or NO_3 oxidation of catechol or methyl catechol under medium NO_x conditions.[27] The relative abundances of 4-nitrocatechol and methyl nitrocatechol (t_R 8.82) were determined to be a factor of 62 and 266 times more significant in the winter sample, respectively. Further, methyl nitrocatechol at t_R 7.15 was not detected in the summer sample, suggesting that the differences observed in the winter C_6 to C_8 CHON grouping are the result of biomass burning influences, which are known to be more abundant in winter.[28–33]

Similarly, Figure 2 shows that the C_{16} to C_{20} CHOS grouping is most abundant in the daytime and the summer season. C_{16} to C_{20} CHOS groupings comprise several species that dominate the chemical composition, as shown in Figure

S8. These compounds were not very oxidized (O/C ratio 0.17 to 0.19) and all displayed similar chemical properties, containing three oxygen atoms and four double bonds and differed in their molecular formula by CH_2. The commercial library identified 4-dodecylbenzenesulfonic acid (spectral match >91% confidence), a surfactant mainly used in laundry detergent and commonly produced in a mixture of linear alkylbenzene sulfonates (LAS).[34] An additional five abundant C_{16} to C_{20} CHOS species could also be observed, including $C_{17}H_{28}O_3S$ (MW 312, t_R 22.37, 22.49, and 22.68) and $C_{16}H_{26}O_3S$ (MW 298, t_R 21.25 and 21.53). All of these compounds displayed fragment ions m/z 183, 119, and 80, corresponding to characteristic LAS fragmentation patterns,[35] supporting the tentative identification of 4-dodecylbenzenesulfonic acid by the commercial library. LASs have been observed in rainwater[36] and fog extracts,[37] although due to their low volatility, the observation of these compounds in PM is rare. Nevertheless, it is worth noting that the sampling site was close to a launderette, which was closed during nighttime hours, possibly accounting for the observed decrease in the nighttime LAS abundance and may indicate a new potential daytime source.

The method also tentatively identified several compounds in the PM samples used in the agricultural industry, with spectral matches >82% confidence for an insecticide (omethoate, $C_5H_{12}NO_4PS$, t_R 2.11), a fungicide (triadimefon, $C_{14}H_{16}ClN_3O_2$, t_R 17.30), and an herbicide (acetochlor, $C_{14}H_{20}ClNO_2$, t_R 18.02). Acetochlor and triadimefon are suspected carcinogenic compounds,[38−40] and omethoate is known to be highly toxic to the aquatic environment.[40] Interestingly, these compounds were only observed in the samples collected overnight in the summer season (samples 264 and 274, Table S2) and appear to be products of long-range transportation from air masses outside of the city (see Figure S9). These pollutants have previously been observed in PM[41−44] and are known to correlate with agricultural activities,[44] remaining airborne for several days following spraying,[45] supporting these observations and modeled data.

Analysis of Surface Waters. Finally, we use the non-targeted method for the compositional analysis of surface waters collected from seven different locations across China, India, and Sri Lanka. The sampling locations are shown in Table S3. Non-targeted screening detected between 1503 to 9165 compounds in the individual samples (sum of positive and negative ionization mode, see Table S9). In contrast to ambient PM, targeted screening using the in-house library identified fewer compounds in the surface water samples. A total of 60 standards were identified, 35 of which were quantified. The commercial library however, offered considerable advances, providing tentative molecular identifications for an additional 94 compounds with spectral matches >85% confidence. The concentrations of the targeted compounds in the surface water samples are shown in Table S10. The measured concentrations of octanedioic and nonanedioic acid, identified in all surface water samples, were found to be remarkably comparable with those recently reported in the River Rhone, France,[46] ranging from 0.11 to 0.49 and 0.16 to 0.70 $\mu g\ L^{-1}$, respectively.

Here, we use the chemical groupings commonly used for the compositional analysis of dissolved organic matter (DOM) but also utilize the strengths of the non-targeted method, grouping all tentatively identified compounds with spectral matches >85% confidence (assigned by the commercial library) by their potential pollutant sources, as shown in Table S11. The commercial library identified several harmful pollutants in the surface water samples, including carcinogenic industrial and agricultural chemicals (tributyl phosphate and carbendazim), active ingredients in pharmaceutical medication (e.g., anaesthetic, analgesic, antipsychotics), stimulants (caffeine; high toxicity risk for aquatic organisms[47]), potential illicit drugs (methamphetamine), and personal care products (N,N-diethyl-m-toluamide, DEET) among others. The identification of these pollutants in surface waters is not uncommon.[47−49] For example, DEET is the primary ingredient used in insect repellents[50] and an abundant and frequently detected pollutant in surface waters.[51−53] DEET was detected in three surface water samples (S1−S3, see Table S3) and was most prominent in the industrial and wastewater effluent samples collected in Sri Lanka, representing >12% of the total sample abundance. In fact, DEET accounted for >45% of the most abundant DOM chemical grouping (i.e., highly unsaturated CHON) in the wastewater effluent sample, as shown in Figure S10, compositional information that would not have been observed without the use of non-targeted screening.

Hospital and urban effluents are the main sources of pharmaceuticals, disinfectants, and detergents in surface waters.[51,54] These pollutants are typically released into the wastewater network, undergoing treatment before being discharged into surface waters, although their removal is known to be inefficient.[49,55] This is exacerbated in developing countries, where limited facilities can result in the discharge of wastewaters into the environment without prior treatment[56] The non-targeted method can identify point-source pollution, providing holistic information on compositional changes detected in samples collected from different geographical locations, as shown in Figure 3 for the River Nag, India. Here, it can be observed that the abundance of several pollutant groupings, including tobacco, stimulants, pharmaceuticals, and industrial chemicals, increased downstream of a major urban area (S1 to S2) and then decreased further downstream after dilution from the confluence of a separate less polluted river (Pili river,[57] S3), following expected trends for potential effluent discharge in the urban area. This was particularly evident for the pharmaceutical grouping, which increased in abundance by a factor of three downstream of two major hospitals, a pollution source previously observed in developing countries.[58] Nagpur produces over 450 million liters of wastewater per day, but less than one-fifth undergoes treatment before being discharged into the River Nag,[57] potentially accounting for these observations.

■ DISCUSSION AND LIMITATIONS

Targeted methods, while suited to regulatory activities, provide limited compositional information and rarely capture the heterogeneity of the natural environment. Automated non-targeted screening tools offer considerable advances within the environmental sciences, but these methods (as shown here) are not infallible. It is imperative that non-targeted methods are rigorously tested to provide a fundamental understanding of their potential limitations, especially as their use increases within the environmental sciences. The developed method, at a minimum, provides a level 5 confidence in molecular assignment (i.e., exact mass) as defined by Schymanski et al. (2014). However, for the vast majority of compounds, a level 3 identification confidence is achieved, providing unambiguous molecular formula assignment and MS^2 data, allowing the
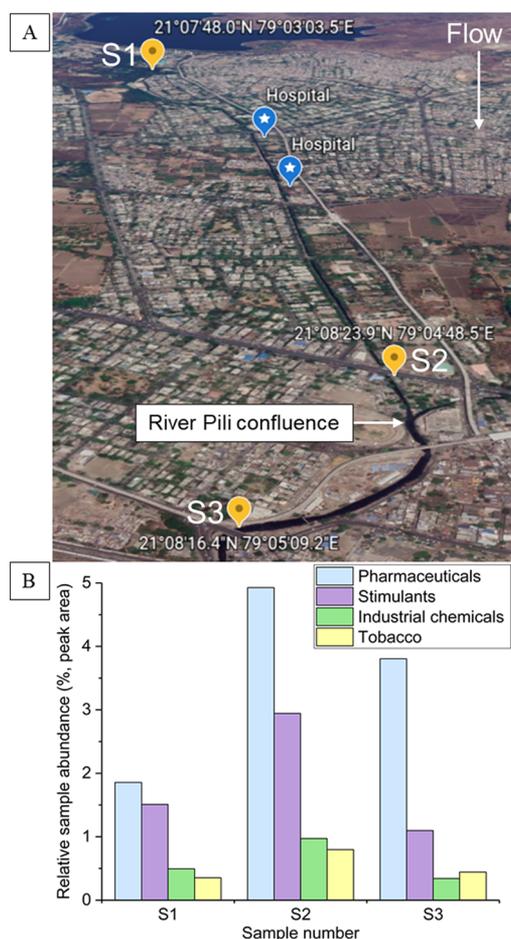
Figure 3. Surface water sampling locations in the River Nag, India (A) and the detected pollutants in each sample with spectral matches >85% confidence (assigned by the commercial mass spectral library, mzCloud) grouped by their potential sources (see Analysis of Surface Waters and Table S11 for further information). The directional flow of the river Nag is shown, along with the confluence of the river Pili. The sample numbers correspond to Table S3. The percentage relative sample abundance was calculated by dividing the measured peak area of each compound by the total sample peak area and multiplying by 100.

compound structure, substituent, or class to be tentatively assigned. The mass spectral libraries provided greater confidence in assignment, including probable structure through MS^2 library matches (level 2) and confirmed structure via targeted screening using authentic standards (level 1, highest identification confidence). We note that there is also further potential to improve the identification of unknowns in the environmental sciences. Commercial MS^2 libraries are rarely used in the environmental domain.[59] We show how the use of the commercial library, mzCloud, can offer advances for non-targeted identification (i.e., rapid identification of probable structure, level 2 confidence[60]), and as this database grows, increased molecular identifications can be anticipated.

One of the main analytical challenges in the analysis of complex matrices is the semiquantitative or quantitative measurement of unknown compounds. Molecular structure can have a considerable impact on ESI efficiencies[61] and sample extraction recoveries.[62] These should be recognized as potential limitations. For example, normalized sample abundance exploited here and commonly used in the

environmental sciences[63−66] does not account for different ESI efficiencies. Large differences in ESI efficiencies of individual compounds may distort or disproportionately affect the normalized abundance of the chemical groupings, particularly where few compounds are detected, or where sample compositions vastly differ. Moreover, to determine the recovery efficiency of the sample extraction procedure, the molecular identity of each compound (out of the thousands detected) must be known. Only then can authentic standards be used to accurately quantify recovery efficiencies, assuming that commercial standards are available (a known difficulty in the environmental domain[67]). Internal standards are also plagued by the same challenge, i.e., often not representative of the sheer number of chemically diverse compounds present in environmental matrices. Here, we used common practice sample extraction procedures for PM (e.g., ref 68 and references therein). We did not however investigate the recovery efficiencies of the authentically identified compounds and therefore recommend that in future work, such analysis is performed to quantify any potential losses and provide insight into the quality of the extraction procedure.

To our knowledge, we provide the most rigorously tested automated non-targeted methodology for the compositional analysis of environmental matrices. While the underlying algorithms in the framework can be further improved (i.e., detection of sodium adducts and use of accurate mass resolution for chromatographic peak detection), the approaches shown offer substantial advances from traditional targeted approaches, providing a more complete assessment of sample compositions while significantly increasing throughput. Data can be analyzed unsupervised on a desktop computer in a few hours, in comparison to months of continuous manual processing.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c08208.

> UPLC-MS and data program method, discussion on the removal of system artifacts, sodium adduct detection, and software notes (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Kelly L. Pereira − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.;* ⓞ orcid.org/0000-0002-4823-4696; Phone: Tel.: +44 (0)1904 321220; Email: kelly.pereira@york.ac.uk

Authors

Martyn W. Ward − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.*

John L. Wilkinson − *Department of Environment and Geography, University of York, York YO10 5NG, U.K.;* ⓞ orcid.org/0000-0002-6993-216X

Jonathan Brett Sallach − *Department of Environment and Geography, University of York, York YO10 5NG, U.K.*

Daniel J. Bryant − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.*

William J. Dixon − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.*

Jacqueline F. Hamilton − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.;* ⦿ orcid.org/0000-0003-0975-4311

Alastair C. Lewis − *Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York YO10 5DD, U.K.*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.0c08208

Author Contributions

K.P. was responsible for the conceptualization, designed and tested the non-targeted and data processing method, analyzed the $PM_{2.5}$ and water samples, and led the production of the manuscript. M.W. wrote the Python code for the data program. J.W., J.B.S., and A.B. collected and prepared the water samples. D.B. and W.D. collected and prepared the PM samples. All authors contributed toward the interpretation and writing of the manuscript.

Notes

The authors declare no competing financial interest.

■ **REFERENCES**

(1) Landrigan, P. J.; Fuller, R.; Acosta, N. J. R.; Adeyi, O.; Arnold, R.; Basu, N.; Baldé, A. B.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J. I.; Breysse, P. N.; Chiles, T.; Mahidol, C.; Coll-Seck, A. M.; Cropper, M. L.; Fobil, J.; Fuster, V.; Greenstone, M.; Haines, A.; Hanrahan, D.; Hunter, D.; Khare, M.; Krupnick, A.; Lanphear, B.; Lohani, B.; Martin, K.; Mathiasen, K. V.; McTeer, M. A.; Murray, C. J. L.; Ndahimananjara, J. D.; Perera, F.; Potočnik, J.; Preker, A. S.; Ramesh, J.; Rockström, J.; Salinas, C.; Samson, L. D.; Sandilya, K.; Sly, P. D.; Smith, K. R.; Steiner, A.; Stewart, R. B.; Suk, W. A.; van Schayck, O. C. P.; Yadama, G. N.; Yumkella, K.; Zhong, M. The Lancet Commission on pollution and health. *The Lancet* **2018**, *391*, 462−512.

(2) Landrigan, P.; Fuller, R.; Acosta, N. J. R.; Adeyi, O.; Arnold, R.; Basu, N.; Baldé, A. B.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J. I.; Breysse, P. N.; Chiles, T.; Mahidol, C.; Coll-Seck, A. M.; Cropper, M. L.; Fobil, J.; Fuster, V.; Greenstone, M.; Haines, A.; Hanrahan, D.; Hunter, D.; Khare, M.; Krupnick, A.; Lanphear, B.; Lohani, B.; Martin, K.; Mathiasen, K. V.; McTeer, M. A.; Murray, C. J. L.; Ndahimananjara, J. D.; Perera, F.; Potočnik, J.; Preker, A. S.; Ramesh, J.; Rockström, J.; Salinas, C.; Samson, L. D.; Sandilya, K.; Sly, P. D.; Smith, K. R.; Steiner, A.; Stewart, R. B.; Suk, W. A.; van Schayck, O. C. P.; Yadama, G. N.; Yumkella, K.; Zhong, M. The Lancet Commission on pollution and health. *The Lancet* **2017**, *391*, 462−512.

(3) Goldstein, A. H.; Galbally, I. E. Known and unexplored organic constituents in the earth's atmosphere. *Environ. Sci. Technol.* **2007**, *41*, 1514−1521.

(4) Houtman, C. J. Emerging contaminants in surface waters and their relevance for the production of drinking water in Europe. *J. Geophys. Res. Atmos.* **2010**, *7*, 271−295.

(5) The European Commision *COMMISSION DIRECTIVE (EU) 2015/1480*; https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015L1480&from=EN

(6) Council, D. E. Amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. *Official Journal of the European Union*, L 226, 29 August 2015, p 226/4−226/11, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2015:226:TOC.

(7) DEFRA *Environmental Protection Act 1990*; 2012.

(8) Fu, Y.; Zhou, Z.; Kong, H.; Lu, X.; Zhao, X.; Chen, Y.; Chen, J.; Wu, Z.; Xu, Z.; Zhao, C.; Xu, G. Nontargeted Screening Method for Illegal Additives Based on Ultrahigh-Performance Liquid Chromatography−High-Resolution Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 8870−8877.

(9) Knolhoff, A. M.; Zweigenbaum, J. A.; Croley, T. R. Nontargeted Screening of Food Matrices: Development of a Chemometric Software Strategy To Identify Unknowns in Liquid Chromatography−Mass Spectrometry Data. *Anal. Chem.* **2016**, *88*, 3617−3623.

(10) Ibáñez, M.; Sancho, J. V.; Hernández, F.; McMillan, D.; Rao, R. Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-light mass spectrometry. *TrAC, Trends Anal. Chem.* **2008**, *27*, 481−489.

(11) Herrera-Lopez, S.; Hernando, M. D.; García-Calvo, E.; Fernández-Alba, A. R.; Ulaszewska, M. M. Simultaneous screening of targeted and non-targeted contaminants using an LC-QTOF-MS system and automated MS/MS library searching. *J. Mass Spectrom.* **2014**, *49*, 878−893.

(12) Nurmi, J.; Pellinen, J.; Rantalainen, A.-L. Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry. *J. Mass Spectrom.* **2012**, *47*, 303−312.

(13) Katajamaa, M.; Orešič, M. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* **2007**, *1158*, 318−328.

(14) Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S. S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; Jahn, D.; Fiehn, O. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. *Aust. J. Chem.* **2017**, *9*, 32.

(15) Fu, Y.; Zhao, C.; Lu, X.; Xu, G. Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography−high resolution mass spectrometry. *TrAC, Trends Anal. Chem.* **2017**, *96*, 89−98.

(16) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf.* **2010**, *11*, 395.

(17) Heald, C. L.; Kroll, J. H.; Jimenez, J. L.; Docherty, K. S.; DeCarlo, P. F.; Aiken, A. C.; Chen, Q.; Martin, S. T.; Farmer, D. K.; Artaxo, P. A simplified description of the evolution of organic aerosol composition in the atmosphere. *Geophys. Res. Lett.* **2010**, *37*, L08803.

(18) Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H.; Mysak, E. R.; Smith, J. D.; Kolb, C. E.; Worsnop, D. R. Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nat. Chem.* **2011**, *3*, 133−139.

(19) Melendez-Perez, J. J.; Martínez-Mejia, M. J.; Eberlin, M. N. A reformulated aromaticity index equation under consideration for non-aromatic and non-condensed aromatic cyclic carbonyl compounds. *Org. Geochem.* **2016**, *95*, 29−33.

(20) Stenson, A. C.; Marshall, A. G.; Cooper, W. T. Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Anal. Chem.* **2003**, *75*, 1275−1284.

(21) Hertkorn, N.; Benner, R.; Frommberger, M.; Schmitt-Kopplin, P.; Witt, M.; Kaiser, K.; Kettrup, A.; Hedges, J. I. Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **2006**, *70*, 2990−3010.

(22) O'Brien, R. E.; Laskin, A.; Laskin, J.; Rubitschun, C. L.; Surratt, J. D.; Goldstein, A. H. Molecular characterization of S- and N-

containing organic constituents in ambient aerosols by negative ion mode high-resolution Nanospray Desorption Electrospray Ionization Mass Spectrometry: CalNex 2010 field study. *J. Geophys. Res.: Atmos.* **2014**, *119*, 12,706−12,720.

(23) Bryant, D. J.; Dixon, W. J.; Hopkins, J. R.; Dunmore, R. E.; Pereira, K. L.; Shaw, M.; Squires, F. A.; Bannan, T. J.; Mehra, A.; Worrall, S. D.; Bacak, A.; Coe, H.; Percival, C. J.; Whalley, L. K.; Heard, D. E.; Slater, E. J.; Ouyang, B.; Cui, T.; Surratt, J. D.; Liu, D.; Shi, Z.; Harrison, R.; Sun, Y.; Xu, W.; Lewis, A. C.; Lee, J. D.; Rickard, A. R.; Hamilton, J. F. Strong anthropogenic control of secondary organic aerosol formation from isoprene in Beijing. *Atmos. Chem. Phys.* **2020**, *20*, 7531−7552.

(24) Wilkinson, J. L.; Boxall, A. B. A.; Kolpin, D. W. A Novel Method to Characterise Levels of Pharmaceutical Pollution in Large-Scale Aquatic Monitoring Campaigns. *Appl. Sci.* **2019**, *9*, 1368.

(25) Pereira, K. L.; Hamilton, J. F.; Rickard, A. R.; Bloss, W. J.; Alam, M. S.; Camredon, M.; Ward, M. W.; Wyche, K. P.; Muñoz, A.; Vera, T.; Vázquez, M.; Borrás, E.; Ródenas, M. Insights into the Formation and Evolution of Individual Compounds in the Particulate Phase during Aromatic Photo-Oxidation. *Environ. Sci. Technol.* **2015**, *49*, 13168−13178.

(26) Kitanovski, Z.; Grgić, I.; Yasmeen, F.; Claeys, M.; Čusak, A. Development of a liquid chromatographic method based on ultraviolet-visible and electrospray ionization mass spectrometric detection for the identification of nitrocatechols and related tracers in biomass burning atmospheric organic aerosol. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 793−804.

(27) Finewax, Z.; de Gouw, J. A.; Ziemann, P. J. Identification and Quantification of 4-Nitrocatechol Formed from OH and NO$_3$ Radical-Initiated Reactions of Catechol in Air in the Presence of NO$_x$: Implications for Secondary Organic Aerosol Formation from Biomass Burning. *Environ. Sci. Technol.* **2018**, *52*, 1981−1989.

(28) Cheng, Y.; Engling, G.; He, K. B.; Duan, F. K.; Ma, Y. L.; Du, Z. Y.; Liu, J. M.; Zheng, M.; Weber, R. J. Biomass burning contribution to Beijing aerosol. *Atmos. Chem. Phys.* **2013**, *13*, 7765−7781.

(29) Salvador, C. M. G.; Tang, R.; Priestley, M.; Li, L.; Tsiligiannis, E.; Le Breton, M.; Zhu, W.; Zeng, L.; Wang, H.; Yu, Y.; Hu, M.; Guo, S.; Hallquist, M. Ambient Nitro-Aromatic Compounds − Biomass Burning versus Secondary Formation in rural China. *Atmos. Chem. Phys.* **2021**, *21*, 1389−1406.

(30) Kahnt, A.; Behrouzi, S.; Vermeylen, R.; Safi Shalamzari, M.; Vercauteren, J.; Roekens, E.; Claeys, M.; Maenhaut, W. One-year study of nitro-organic compounds and their relation to wood burning in PM10 aerosol from a rural site in Belgium. *Atmos. Environ.* **2013**, *81*, 561−568.

(31) Laskin, A.; Laskin, J.; Nizkorodov, S. A. Chemistry of Atmospheric Brown Carbon. *Chem. Rev.* **2015**, *115*, 4335−4382.

(32) Desyaterik, Y.; Sun, Y.; Shen, X.; Lee, T.; Wang, X.; Wang, T.; Collett, J. L., Jr. Speciation of "brown" carbon in cloud water impacted by agricultural biomass burning in eastern China. *J. Geophys. Res.: Atmos.* **2013**, *118*, 7389−7399.

(33) Kitanovski, Z.; Grgić, I.; Vermeylen, R.; Claeys, M.; Maenhaut, W. Liquid chromatography tandem mass spectrometry method for characterization of monoaromatic nitro-compounds in atmospheric particulate matter. *J. Chromatogr. A* **2012**, *1268*, 35−43.

(34) National Library of Medicine *4-Dodecylbenzenesulfonic acid*; https://pubchem.ncbi.nlm.nih.gov/compound/4-Dodecylbenzenesulfonic-acid((accessed on Sept. 25, 2019)),

(35) LeClair, J. P. Structural characterization of water-soluble atmospheric organic matter by ultrahigh-resolution mass spectrometry. Michigan Technological University, 2011.

(36) Altieri, K. E.; Turpin, B. J.; Seitzinger, S. P. Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry. *Atmos. Chem. Phys.* **2009**, *9*, 2533−2542.

(37) Mazzoleni, L. R.; Ehrmann, B. M.; Shen, X.; Marshall, A. G.; Collett, J. L., Jr. Water-Soluble Atmospheric Organic Matter in Fog: Exact Masses and Chemical Formula Identification by Ultrahigh-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Environ. Sci. Technol.* **2010**, *44*, 3690−3697.

(38) Lerro, C. C.; Koutros, S.; Andreotti, G.; Hines, C. J.; Blair, A.; Lubin, J.; Ma, X.; Zhang, Y.; Beane Freeman, L. E. Use of acetochlor and cancer incidence in the Agricultural Health Study. *Int. J. Cancer* **2015**, *137*, 1167−1175.

(39) ECHA *Proposing harmonised classification and labelling at EU level of Acetochlor (ISO); 2-chloro-N-(ethoxymethyl)-N- (2-ethyl-6-methylphenyl)acetamide*; 2014.

(40) Lewis, K. A.; Tzilivakis, J.; Warner, D. J.; Green, A. An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment: An International Journal* **2016**, *22*, 1050−1064.

(41) Coscollà, C.; Colin, P.; Yahyaoui, A.; Petrique, O.; Yusà, V.; Mellouki, A.; Pastor, A. Occurrence of currently used pesticides in ambient air of Centre Region (France). *Atmos. Environ.* **2010**, *44*, 3915−3925.

(42) Coscollà, C.; Hart, E.; Pastor, A.; Yusà, V. LC-MS characterization of contemporary pesticides in PM10 of Valencia Region, Spain. *Atmos. Environ.* **2013**, *77*, 394−403.

(43) Coscollà, C.; Yahyaoui, A.; Colin, P.; Robin, C.; Martinon, L.; Val, S.; Baeza-Squiban, A.; Mellouki, A.; Yusà, V. Particle size distributions of currently used pesticides in a rural atmosphere of France. *Atmos. Environ.* **2013**, *81*, 32−38.

(44) Coscollà, C.; Muñoz, A.; Borrás, E.; Vera, T.; Ródenas, M.; Yusà, V. Particle size distributions of currently used pesticides in ambient air of an agricultural Mediterranean area. *Atmos. Environ.* **2014**, *95*, 29−35.

(45) Avino, P.; Cinelli, G.; Notardonato, I.; Russo, M. V. Investigation on the Behavior of Pesticides in Atmosphere. *Aerosol Air Qual. Res.* **2011**, *11*, 783−790.

(46) Sempéré, R.; Charrière, B.; Castro-Jiménez, J.; Kawamura, K.; Panagiotopoulos, C. Occurrence of α, ω-dicarboxylic acids and ω-oxoacids in surface waters of the Rhone River and fluxes into the Mediterranean Sea. *Prog. Oceanogr.* **2018**, *163*, 136−146.

(47) Deo, R. P. Pharmaceuticals in the Surface Water of the USA: A Review. *Curr. Environ. Health Rep.* **2014**, *1*, 113−122.

(48) Kolpin, D. W.; Furlong, E. T.; Meyer, M. T.; Thurman, E. M.; Zaugg, S. D.; Barber, L. B.; Buxton, H. T. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999−2000: A National Reconnaissance. *Environ. Sci. Technol.* **2002**, *36*, 1202−1211.

(49) Benotti, M. J.; Brownawell, B. J. Distributions of pharmaceuticals in an urban estuary during both dry- and wet-weather conditions. *Environ. Sci. Technol.* **2007**, *41*, 5795−5802.

(50) Kitchen, L. W.; Lawrence, K. L.; Coleman, R. E. The role of the United States military in the development of vector control products, including insect repellents, insecticides, and bed nets. *J. Vector Ecol.* **2009**, *34*, 50−61.

(51) dos Santos, M. M.; Hoppe-Jones, C.; Snyder, S. A. DEET occurrence in wastewaters: Seasonal, spatial and diurnal variability - mismatches between consumption data and environmental detection. *Environ. Int.* **2019**, *132*, 105038.

(52) Merel, S.; Snyder, S. A. Critical assessment of the ubiquitous occurrence and fate of the insect repellent N,N-diethyl-m-toluamide in water. *Environ. Int.* **2016**, *96*, 98−117.

(53) Kolpin, D. W.; Furlong, E. T.; Meyer, M. T.; Thurman, E. M.; Zaugg, S. D.; Barber, L. B.; Buxton, H. T. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999-2000: A national reconnaissance. *Environ. Sci. Technol.* **2002**, *36*, 1202−1211.

(54) Orias, F.; Perrodin, Y. Characterisation of the ecotoxicity of hospital effluents: A review. *Sci. Total Environ.* **2013**, *454-455*, 250−276.

(55) Shraim, A.; Diab, A.; Alsuhaimi, A.; Niazy, E.; Metwally, M.; Amad, M.; Sioud, S.; Dawoud, A. Analysis of some pharmaceuticals in municipal wastewater of Almadinah Almunawarah. *Arabian J. Chem.* **2017**, *10*, S719−S729.

(56) Nations, U. *UN World Water Development Report, Wastewater: The Untapped Resource*; 2017.

(57) Archana, G.; Dhodapkar, R.; Kumar, A. Offline solid-phase extraction for preconcentration of pharmaceuticals and personal care products in environmental water and their simultaneous determination using the reversed phase high-performance liquid chromatography method. *Environ. Monit. Assess.* **2016**, *188*, 512.

(58) Guruge, K. S.; Goswami, P.; Tanoue, R.; Nomiyama, K.; Wijesekara, R. G. S.; Dharmaratne, T. S. First nationwide investigation and environmental risk assessment of 72 pharmaceuticals and personal care products from Sri Lankan surface waterways. *Sci. Total Environ.* **2019**, *690*, 683−695.

(59) Plassmann, M. M.; Fischer, S.; Benskin, J. P. Nontarget Time Trend Screening in Human Blood. *Environ. Sci. Technol. Lett.* **2018**, *5*, 335−340.

(60) 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*, 2097−2098.

(61) Oss, M.; Kruve, A.; Herodes, K.; Leito, I. Electrospray Ionization Efficiency Scale of Organic Compounds. *Anal. Chem.* **2010**, *82*, 2865−2872.

(62) Priego-Capote, F.; Luque de Castro, M. D. Analytical uses of ultrasound I. Sample preparation. *Anal. Chem.* **2004**, *23*, 644−653.

(63) Sato, K.; Fujitani, Y.; Inomata, S.; Morino, Y.; Tanabe, K.; Ramasamy, S.; Hikida, T.; Shimono, A.; Takami, A.; Fushimi, A.; Kondo, Y.; Imamura, T.; Tanimoto, H.; Sugata, S. Studying volatility from composition, dilution, and heating measurements of secondary organic aerosols formed during α-pinene ozonolysis. *Atmos. Chem. Phys.* **2018**, *18*, 5455−5466.

(64) Gao, S.; Surratt, J. D.; Knipping, E. M.; Edgerton, E. S.; Shahgholi, M.; Seinfeld, J. H. Characterization of polar organic components in fine aerosols in the southeastern United States: Identity, origin, and evolution. *J. Geophys. Res.: Atmos.* **2006**, *111*, D14314.

(65) Ladd, M. P.; Giannone, R. J.; Abraham, P. E.; Wullschleger, S. D.; Hettich, R. L. Evaluation of an untargeted nano-liquid chromatography-mass spectrometry approach to expand coverage of low molecular weight dissolved organic matter in Arctic soil. *Sci. Rep.* **2019**, *9*, 5810.

(66) Deng, Y.; Inomata, S.; Sato, K.; Ramasamy, S.; Morino, Y.; Enami, S.; Tanimoto, H. Temperature and acidity dependence of secondary organic aerosol formation from α-pinene ozonolysis with a compact chamber system. *Atmos. Chem. Phys.* **2021**, *21*, 5983−6003.

(67) Finessi, E.; Lidster, R. T.; Whiting, F.; Elliott, T.; Alfarra, M. R.; McFiggans, G. B.; Hamilton, J. F. Improving the Quantification of Secondary Organic Aerosol Using a Microflow Reactor Coupled to HPLC-MS and NMR to Manufacture Ad Hoc Calibration Standards. *Anal. Chem.* **2014**, *86*, 11238−11245.

(68) Zielinski, A. T.; Campbell, S. J.; Seshia, A. A.; Jones, R. L.; Kalberer, M.; Giorio, C. Compositional Analysis of Adsorbed Organic Aerosol on a Microresonator Mass Sensor. *Aerosol Sci. Eng.* **2018**, *2*, 118−129.