PFΔ*Screen* – An open-source tool for automated PFAS feature prioritization in non-target HRMS data

Jonathan Zweigle,[+,*] Boris Bugsel,[+] Joel Fabregat-Palau,[‖] Christian Zwiener[+,*]

[+]Environmental Analytical Chemistry, Department of Geosciences, University of Tübingen, Schnarrenbergstraße 94-96, 72076 Tübingen, Germany

[‖]Hydrogeochemistry, Department of Geosciences, University of Tübingen, Schnarrenbergstraße 94-96, 72076 Tübingen, Germany

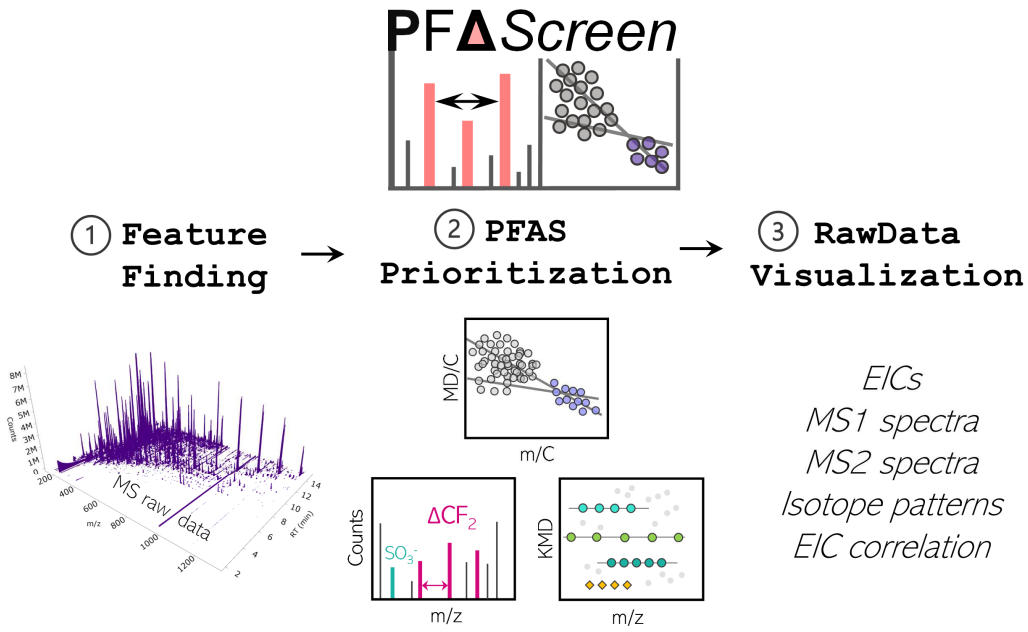
[*]*Corresponding authors*

Abstract

Per- and polyfluoroalkyl substances (PFAS) are a huge group of anthropogenic chemicals with unique properties that are used in countless products and applications. Due to the high stability of their C–F bonds, PFAS or their transformation products (TPs) are persistent in the environment, leading to ubiquitous detection in various samples worldwide. Since PFAS are industrial chemicals, the availability of authentic PFAS reference standards is limited, making non-target screening (NTS) approaches based on high-resolution mass spectrometry (HRMS) necessary for a more comprehensive characterization. NTS usually is a time-consuming process, since only a small fraction of the detected chemicals can be identified. Therefore, efficient prioritization of relevant HRMS signals is one of the most crucial steps. We developed PFΔ*Screen*, a Python-based open-source tool with a simple graphical user interface (GUI) to perform efficient feature prioritization by several PFAS specific techniques such as the highly promising MD/C-m/C approach, Kendrick mass defect analysis, diagnostic fragments (MS^2), fragment mass differences (MS^2) and suspect screening. Feature detection from vendor-independent MS raw data (mzML, data-dependent acquisition) is performed via pyOpenMS (or custom feature lists) with subsequent calculations for prioritization and identification of PFAS in both HPLC- and GC-HRMS data. The PFΔ*Screen* workflow is presented on four PFAS-contaminated agricultural soil samples from south-western Germany. Over 15 classes of PFAS (more than 80 single compounds with several isomers) could be identified, including four novel classes, potentially TPs of the precursors fluorotelomer mercapto alkyl phosphates (FTMAPs). PFΔ*Screen* can be used within the Python environment and is easily automatically installable and executable on Windows. Its source code is freely available on GitHub (https://github.com/JonZwe/PFAScreen).

Keywords: PFAS, non-target screening, feature prioritization, HRMS, open-source software, mass defect, MD/C-m/C, KMD

Graphical abstract:



Introduction

Per- and polyfluoroalkyl substances (PFAS) are a large group of anthropogenic chemicals characterized by containing multiple C-F bonds [1,2]. Due to their unique properties, they are used in a wide array of daily products and different industrial applications [3]. Their high chemical resistance and water- and oil repellency lead to the production of PFAS with a variety of different chemistries [4]. Due to the high stability of C-F bonds, the perfluoroalkyl chains of PFAS exhibit an intrinsic persistence that leads to a worldwide distribution of PFAS and their terminal transformation products (TPs) such as perfluoroalkyl acids (PFAAs) which were extensively produced and used in the past [5-8]. Nowadays, the number of known PFAS ranges from thousands to millions, depending on the definition and source of information. According to the updated OECD definition, all chemicals containing a CF_3- or isolated CF_2-group are considered PFAS, which has increased the number of PFAS considerably [9,10]. Global regulatory efforts restricted the production of selected longer-chain PFAAs such as e.g., perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) due to their persistence, bioaccumulation potential, and adverse effects on humans and the environment [11]. This resulted in the production of replacement compounds of rather similar persistence, increasing the number of different PFAS on the global market that are also eventually emitted into the environment [12]. Therefore, PFAS are considered to be regulated as a chemical class in the European Union in the future [13].

Several studies have shown that considerable fractions of organically bound fluorine (e.g., extractable organic fluorine) in environmental and human samples cannot be explained sufficiently by routinely analyzed PFAS (target screening), which usually include less than 50 analytes [14-17]. Since almost no fluorinated organic compounds occur naturally, unknown fractions of organically bound fluorine are clear indications of anthropogenic chemicals [18].

Due to the sheer number of different PFAS that transform into an even larger number of unknown TPs, a comprehensive use of authentic reference standards is usually not possible and most likely will not be soon [19,20]. The fact that PFAS are industrial chemicals that often underlie the trade secrets even complicates the availability of standards. Therefore, non-target screening (NTS) based on high-resolution mass spectrometry (HRMS) is necessary for a more comprehensive characterization of PFAS [21,22]. Several studies have shown that target analysis is insufficient to capture PFAS present in complex samples, which can easily result in the overlooking of important compounds even when present in high concentrations [23]. NTS-approaches led to the identification of more than 750 novel PFAS in various samples in the past worldwide, showing their high relevance in analytical approaches [24,22]. Since NTS is typically a time-consuming and often partially manual process, efficient prioritization techniques are needed to separate detected matrix components from the analytes of interest (often a data reduction from ~5000 detected compounds to 10-100 identified analytes or even less) [25].

The intrinsic properties of PFAS (with a certain fluorine percentage) allow the use of several techniques for their prioritization [21,26]: The chemical mass defect (MD) of PFAS is typically lower ($MD_F = -1.6\times10^{-3}$ Da) than the one of hydrocarbons ($MD_H = +7.8\times10^{-3}$ Da) and has been used to remove detected features outside a predefined MD range (e.g., -0.25 Da to +0.1 Da) [27-29]. However, this range is not fixed, and depending on the structure, it is important to know that hydrocarbons of higher mass that exceed a MD of +0.75 Da can also fall into the same range. Similarly, polyfluorinated PFAS with a high H-content may bear a positive MD exceeding +0.1 Da. Recently, a promising approach based on the MD normalized to the carbon number (MD/C) vs. the mass normalized to the carbon number (m/C) was proposed to separate PFAS much more efficiently from other hydrocarbon features in HRMS data which was further systematically evaluated for ~200,000 PFAS from chemical databases [30,26]. The carbon number can be easily estimated for all HRMS features by using the

relative abundance of the M+1 isotope (^{13}C). PFAS have a much higher m/C when their mass is dominated by fluorine (e.g., m/C ~ 50), while hydrocarbons of similar mass are dominated by carbon (m/C ~ 14), allowing a convenient separation. Details on the MD/C-m/C approach are summarized in Zweigle et al. 2023 [26]. Especially, the m/C dimension can be used to remove large fractions of non-PFAS features when applied appropriately. This is illustrated in Fig. 1 where we plotted a 2D histogram of the MD/C-m/C locations of over 50000 features from previous HRMS measurements of PFAS-contaminated soils and grease-repelling papers, where a clear separation of potentially highly fluorinated compounds is observed (region around m/C ≈ 40, MD/C = -0.002). It is important to note, however, that the MD/C-m/C separation works better the higher the percentage of fluorine in a molecule is, with an accordingly, higher F/C and a lower H/F ratio [26]. Like the MD, the MD/C-m/C approach cannot separate, for instance, hydrocarbons with one or two CF_3 groups from other hydrocarbons.

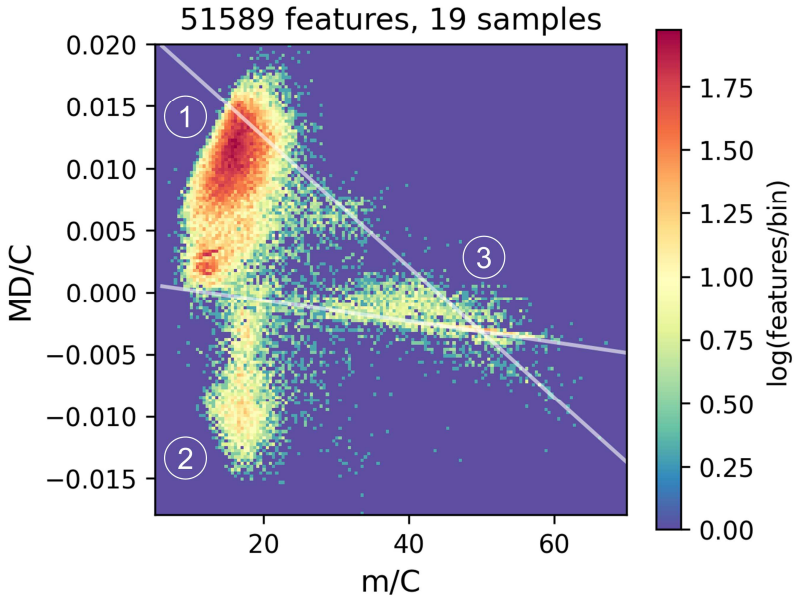


Fig. 1: 2D histogram of the number of compounds (log scale) (compound density) in the MD/C-m/C plot of 19 measured samples used from several paper and soil extracts, standards, and blanks (19 samples with 51589 features from [23,31,32]). Hydrocarbon features are located usually below m/C of 25 with a clearly positive MD/C (position 1), while at a certain C number the MD exceeds +0.5 Da yielding a position of a mathematical negative MD/C (position 2). Highly fluorinated compounds or compounds with other heavy heteroatoms are strongly shifted to higher m/C values (position 3). It becomes obvious that even with these high numbers of features in several samples from several different matrices, potential PFAS features with a certain fraction of fluorine within the molecule are efficiently separated from most matrix components The grey lines mark the CH_xF_{2-x}-line ($0 \leq x \leq 2$) and the CF_x-line ($0 \leq x \leq 2$) (for details on the MD/C-m/C plot see Zweigle et al. [26]).

Besides the MD and MD/C-m/C-approach, the Kendrick mass defect (KMD) analysis to detect homologous series of PFAS (e.g., with CF_2 or CF_2O as repeating units) is of great relevance since it allows the grouping of structurally related PFAS, simplifying their identification [27,33]. In the MS^2 data, lists of PFAS-specific diagnostic fragments (DFs) as well as fragment mass differences and neutral losses can be used to prioritize fragmentation spectra [28,34,31]. These techniques are often combined with suspect screening by matching accurate mass (or further evidence) with PFAS lists [22,35].

KMD, DFs, fragment mass differences, and especially suspect screening with large lists (e.g., PFASMASTER, gathering over 12,000 compounds [36]) in combination with complex samples (thousands of features) are prone to a high number of false-positive detections (depending on mass tolerance) that often need to be excluded manually, which is a time-consuming process. Even with extremely high mass resolution, naturally occurring compounds can still mimic certain PFAS-specific repeating units such as CF_2, complicating KMD analysis and making retention time shifts a necessary criterion [37]. Therefore, if the number of features can be preliminarily reduced by the MD/C-m/C approach before applying those techniques, a faster and more accurate NTS-workflow can be performed, decreasing both computational and manual effort regarding the further inspection of the features.

To facilitate the non-targeted screening of PFAS in complex samples, we developed PFΔ*Screen*, an open-source Python-based software tool with a simple graphical user interface (GUI) that combines the discussed techniques to efficiently prioritize PFAS in LC- or GC-HRMS data. PFΔ*Screen* can be applied vendor-independently either on mass spectrometric raw data (mzML, automated feature finding via pyOpenMS) or on custom feature lists (external feature finding by other software tools). The PFΔ*Screen* workflow is here presented by application to four PFAS-contaminated agricultural soil extracts from south-western Germany (Rastatt case [38,27]), where several PFAS classes, including novel PFAS, were identified. The advantages of the combined workflow are discussed in detail. The source code

is available via GitHub and can be easily automatically installed and executed via batch files on Windows within the Python environment.

Materials and Methods

PFΔ*Screen* workflow

PFΔ*Screen* is a fully automated tool for detection and prioritization of potential PFAS features (LC- or GC-HRMS) in raw mass spectrometric data written in Python (3.9.13) (Fig. 2). PFΔ*Screen* is structured in several individual Python functions that are executed from one main file that allows data and parameter input via a simple GUI programmed with the tkinter library (Fig. S1). It can easily be automatically installed and executed on Windows using batch files. Detailed instructions on installation and functionality are provided in the SI. Input MS raw data can be converted vendor-independently from data-dependent acquisition (ddMS^2) files into the mzML data format (.mzML) by using the MSConvert software from ProteoWizard [39,40]. Only mzML files with centroided spectra and one collision energy (CE) should be used. If profile data was acquired and MS^2 spectra from several different CEs per precursor m/z are present, the peak picking (for centroiding) and subset functions (to keep only one desired CE) from MSConvert can be used to generate the correct mzML input files.

In the following, the three main functionalities of PFΔ*Screen* are explained in the same order as they can be executed in the GUI (Fig. 2 and Fig. S1).

FeatureFinding.

The first step usually performed in NTS is detection of features in the MS raw data characterized by chromatographic peak shapes of coeluting isotopes, resulting in a list of m/z, retention time (RT) and peak area. This task is performed with pyOpenMS, a Python interface to the C++ OpenMS library [41-45]. For feature detection, the FeatureFinderMetabo algorithm is used, which is designed for metabolites and small molecules [46-48]. Three parameters

(mass error (ppm), intensity threshold and an isotope model for more accurate detection of coeluting isotopologues) can be specified. The most important parameter is the intensity threshold, which is highly dependent on the instrument used, sample, and the underlying NTS question. After feature finding in the MS^1 data, MS^2 spectra can be aligned to their respective precursors by specifying an m/z- and RT-tolerance. Only one unique MS^2 spectrum with the highest precursor intensity is assigned to the respective MS^1 precursor.

With PFΔ*Screen,* a single sample with a corresponding (optional) blank can be processed at a time. Blank correction is performed by setting an m/z- and RT-tolerance as well as a fold change with the desired increase of abundance in the sample compared to the blank. Features appearing in both sample and blank within the specified criteria are removed from the dataset. After preprocessing, the raw data is ready for specific PFAS prioritization. If feature finding by an external software is desired (e.g., vendor software), the following steps can also be performed by loading a feature table (.xlsx, that requires m/z, RT, and intensities of the [M] and [M+1] isotopes) into PFΔ*Screen* without feature detection via OpenMS. However, the raw mzML files are still needed to assign MS^2 data to the features in the feature table (see **SI**). Besides pyOpenMS, the mass spectrometric Python library Pyteomics is used for selected calculations [49,50].

PFASPrioritization.

The PFAS prioritization workflow is intended in an iterative manner: after feature detection, the MD/C-m/C plot should firstly be manually inspected to determine reasonable boundaries to remove most of the detected features (e.g., ~90%) that cannot be PFAS due to their MD/C-m/C locations (depending on the underlying question). After determination of these cutoffs, the PFAS feature prioritization can be executed again focused on a subset of features, which will strongly decrease false-positives in KMD analysis, fragment matching, and suspect screening where the respective parameters can be adjusted accordingly without a strong increase of wrong assignments. Since the execution time of PFΔ*Screen* is usually below one

minute (e.g., for ~4000 spectra per sample), input parameters can easily be varied to test their influence on the outcome. After execution, a folder is generated named after the sample file where important results are saved, including a summary in an Excel sheet which is formatted as a table that can be easily inspected, sorted and subset for a faster overview of the results (Fig. S2). Important plots are saved in the interactive HTML format which can easily be opened in any browser, allowing zooming and data inspection with interactive tooltips (Fig. S3).

In the workflow to prioritize features according to their likelihood of being PFAS, several pieces of evidence are calculated individually for all detected features in the first place. For all MS^1 features, the number of carbon atoms, MD, and both MD/C and m/C-dimensions are determined. To detect homologues series (HS), the KMD (with a predefined repeating unit required; e.g., CF_2) is calculated and corresponding features belonging to a certain HS are aligned by providing a unique HS number (parameters: mass tolerance, minimum number of homologues).

For all MS^2 spectra, fragment mass differences are calculated comprehensively. Therefore, all fragment differences within each MS^2 spectrum are calculated and matched against a predefined list of PFAS typical mass differences (e.g., ΔCF_2, ΔC_2F_4, ΔHF, $\Delta C_{10}H_3F_{17}$, more details can be found in [31]). This allows an efficient detection of fragments indicative for PFAS without prior knowledge on their actual mass [31,23]. Furthermore, a list of typical PFAS diagnostic fragments (DFs, approximately 900 fragments) from literature are automatically matched with all fragmentation spectra (which is easily extendable) [51,52]. Both negative and positive fragments are considered depending on the measurement polarity which can be specified in the GUI. The most important parameter is the MS^2 noise threshold, used to specify the lowest MS^2 intensity to be considered for DF, and mass difference matching. It is of importance to select a suitable instrument-specific threshold as a too low input value may result in a high number of false-positive annotations. Besides a mass tolerance

for fragment matching, a minimal number of positive DFs or mass differences can be specified to flag a MS^2 spectrum as potential hit.

To enhance annotation in the MS^2, fragments that have a defined mass difference to another already annotated fragment (accurate mass match and therefore also a chemical formula) are also annotated by subtraction or addition of the respective mass difference (e.g., ΔC_2F_4) to an annotated chemical formula (e.g., $C_{12}H_5F_{12}O_4S$ + ΔC_2F_4). This allows the calculation of unknown chemical formulas for fragment masses that are not present in the list of DFs (see Fig. S6).

In the third step, suspect screening by accurate mass match (with mass tolerance) can be performed. We used the PFAS NIST suspect list as a template, with extension of other in-house identified PFAS. For suspect screening, three adducts can be chosen which are $[M–H]^-$ for negative polarity, and both $[M+H]^+$ and $[M]^+$ for positive polarity (compounds such as betaines present in various AFFF formulations are often detected as M^+ ions) [53].

RawDataVisualization

After feature finding or the complete workflow, the MS raw data can be directly visualized via the PFΔ*Screen* GUI (Fig. 2 and S4).

EIC extractor: Extracted ion chromatograms (EICs) can be generated by accurate m/z (e.g., from the Excel results file) and inspected in an external window. Several masses can be extracted together (comma separated) to investigate coelution or RT-shifts. To verify the systematic RT-shifts of detected HS, a repeating unit can be specified (e.g., CF_2) and *n* EICs are extracted at once (Fig. 2 and S4), allowing fast checking for reasonable of peak shapes and elution order of suspected masses.

MS^1 extractor: To visualize single MS^1 spectra, a certain RT of interest can be specified. Theoretical isotope patterns of chemical formulas from suspect hits can then be plotted on top of the experimental MS^1 isotope pattern (Fig. 2 and S4).

MS² extractor: MS² spectra can also be directly accessed via the GUI by inputting the accurate m/z value. If DFs and fragment mass differences were detected, they are displayed within the respective MS² spectrum (Fig. S4).

EIC correlator: To detect potential in-source fragments (e.g., $[M-HF]^-$) or adducts (e.g., $[M+Br]^-$ or $[M+Acetate]^-$) by coelution correlation, an m/z of interest can be specified and all detected features within a certain RT-range are correlated (EICs) and only highly correlating ions can be visualized (e.g., correlation of $R^2 > 0.95$). This can greatly enhance understanding of ionization processes and helps to find related ions that were not grouped during feature detection (more detailed explanation in the Results & Discussion section, Fig. 6 and S9).

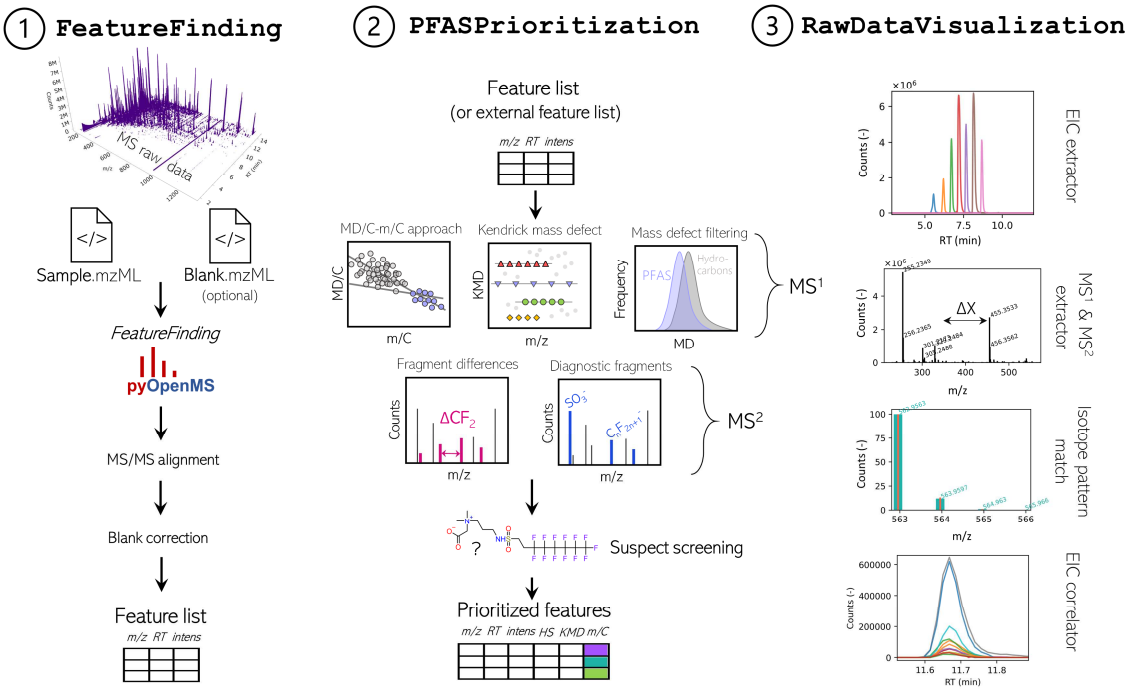


Fig. 2: Schematic overview of the PFΔ*Screen* workflow in the structure of the GUI (Fig. S1). The FeatureFinding tab (1) allows detection of feature via pyOpenMS in MS raw data followed by MS² alignment and blank correction resulting in a feature list for a sample of interest. PFAS feature prioritization (2) includes techniques such as the MD/C-m/C approach, KMD analysis, fragment matching, and fragment mass differences which generates a strongly reduced feature list of potential PFAS. The data from this list can be visualized and verified by the RawDataVisualization tool (3) together with other output file such as interactive HTML plots which allows efficient NTS (Fig. S3-S5).

Soil collection and extraction

To present the feature prioritization procedure via PFΔ*Screen*, four different PFAS-contaminated composite agricultural topsoil samples from Rastatt (R1 & R2) and Mannheim

(M1, M2) regions (Germany) were extracted and measured by HPLC-QTOF-MS (see sampling details and soil physicochemical properties in the SI (S3)). The R1, R2, S1, and S2 soil names correspond to soils B, A, D, and H from Röhler et al (2023), respectively [54]. Agricultural fields in these regions were subjected to contaminated paper sludge in the past and found to be highly contaminated with several PFAS classes [27,32,54]. Information on all chemicals used can be found in SI (S4). Soil extraction was adapted from existing procedures [27]. Briefly, five g of dried soil (40 °C) were weighed in 50 mL polypropylene (PP) tubes and combined with 10 mL of methanol (MeOH). The suspension was sonicated for one hour and overhead shaken for 16 hours. After centrifugation (10 min @ 4000 rcf), the supernatant was transferred into a 20 mL glass vessel, and extraction was repeated. The combined extracts (20 mL) were evaporated under a gentle stream of N_2 until dryness at 40 °C and reconstituted in one mL of MeOH, sonicated for 10 min and thoroughly vortexed for one min. In a last step, the enriched extract was filtered through a 0.2 µm regenerated cellulose syringe filter, transferred into PP HPLC vials, and stored in the fridge (4°C) until analysis. As quality control, an extraction blank following the identical extraction procedure but without adding any soil was prepared to account for background contamination.

LC-HRMS measurements and data acquisition

Soil extracts were analyzed with an Agilent 1260 Infinity HPLC system (Poroshell 120 EC-C_{18} column; 2.1 mm × 100 mm; 2.7 µm particles at 40 °C) at a flow rate of 0.3 mL/min coupled to an Agilent 6550 QTOF-mass spectrometer. For compound separation, a 23 min gradient program was used (A: 95/5 H_2O/MeOH + 2 mM NH_4Ac; B: 5/95 H_2O/MeOH + 2 mM NH_4Ac) and both negative and positive measurements were performed (details in Table S1-S2). Data acquisition was performed in the data-dependent mode (ddMS^2) using 3 scans/s (MS^1 range: m/z 100−1700 and MS^2 range m/z 70−1700) with a static exclusion list (resulting from prior MeOH blank injections) to avoid fragmentation of background signals. Furthermore, a rolling

exclusion list was used to iteratively exclude previously triggered precursor masses from previous measurements (three injections) of the same sample to maximize the MS^2 coverage. The threshold for precursor selection was set to 1000 counts, and each precursor was excluded for 0.5 min after collection of three MS^2 spectra. For collision induced dissociation, a linear m/z-dependent collision energy (CE) according to the following equation was used: $CE(m/z) = 3\frac{m/z}{100}+15$ eV. To prevent sample cross contamination, a three-fold needle wash in MeOH was performed in-between each injection. Each measurement sequence included several blanks and quality controls (PFAS reference standard mixture) to monitor instrument drift.

Results and Discussion

PFAS prioritization and identification with PFΔ*Screen* is aimed to be performed in an iterative process. This means that the program is executed multiple times allowing parameter adjustment to generate reasonable results. PFΔ*Screen* runtimes are usually below one minute (e.g., for ~4000 spectra per sample) for the whole workflow. When changing specific input parameters (e.g., tolerances, thresholds, mass differences etc.), their effect on the output can directly be observed. In this way, input parameters can be conveniently adjusted depending on end-user needs and sample types. After feature detection, blank correction, and a short inspection of the results, the data can be reduced by the MD/C-m/C approach by setting an appropriate m/C cutoff value. Subsequent KMD analysis, fragment mass differences, DF matching and suspect screening then result in a detailed table of a manageable size.

To demonstrate the PFΔ*Screen* workflow, it was here applied to four contaminated agricultural topsoils. We started the iterative identification process with the soil extract of M1. After data preprocessing and application of prioritization techniques, the identified PFAS (including adducts and in-source fragments) were manually added to the suspect list, and the same workflow was applied to the next soil sample. In the following, the whole workflow starting from data reduction to final identification is discussed in detail.

Data preprocessing

After data-dependent acquisition (DDA), the raw MS data (.d files, Agilent) were converted into mzML with MSConvert [39]. For each soil, PFΔ*Screen* was executed individually together with the extraction blank to remove background signals originating from both the extraction procedure and the HPLC system. The mass error for feature detection was set to 10 ppm, the MS^1 intensity threshold was set to 2000 counts and the metabolites (5% RMS) isotope model from OpenMS was used to exclude features with unusual peak shapes of isotopic traces. Peaks reported after feature detection have to have a full width at half

maximum (FWHM) above one second and below 1 minute, and at least two isotopic traces. MS^2 spectra were aligned with a mass tolerance of 5 mDa and an RT tolerance of 0.2 minutes (these tolerances can be verified by an interactive m/z vs. RT plot (Fig **S3a**)). Features detected in both sample and extraction blank that deviated by < 2 mDa at a RT difference of < 0.1 minute and were not at least 5-fold more abundant in the sample were removed. Exemplified on soil M1, 4209 features were detected, that were reduced to 3750 features after blank correction in the ESI^- mode. A total of 1026 out of 2450 acquired MS^2 spectra corresponded to detected features, from which 417 unique spectra remained (~11% MS^2 coverage in first iteration).

Data reduction by m/C and MD/C

After these feature preprocessing steps, the m/C and MD/C dimensions were used for data reduction. When looking at the MD/C-m/C plot of all soils together (containing more than 12,000 features), a clear separation of three groups of compounds can be observed (Fig. **3a**). Most features were located below m/C 30, which are a wide variety of different hydrocarbon molecules. A theoretical molecule exclusively consisting of $(CH_2)_n$-groups would be located at m/C = 14, while for the four soil extracts a clear peak distribution ranging from m/C ≈ 10 - 25 and reaching a maximum around m/C ≈ 16 was observed (Fig. **3c**). The determination of the carbon number strongly depends on the peak picking algorithm, since it is based on robustly integrated EICs from the monoisotopic mass and its corresponding M+1 isotope ($C \approx I_{M+1}/I_M/0.011145$). Therefore, a certain uncertainty should always be expected, which increases with decreasing ion abundance. Nonetheless, ~92% of all detected features are clearly located below m/C = 30 (e.g., humic substances) (Fig. **3c**). Therefore, here a cutoff at m/C = 30 was chosen since PFAS that are dominated by fluorine usually have a higher m/C (e.g., $m/C_{6:2\ diPAP} \approx 49$; $m/C_{PFOA} \approx 51$; $m/C_{6:2\ FTAB} \approx 38$). 6:2 FTAB is an AFFF constituent which already has a considerable fraction of hydrogen ($C_{15}H_{19}F_{13}N_2O_4S$) compared to other

PFAS, while other organic compounds containing less fluorine (compared to hydrogen, high H/F ratio) such as the pharmaceutical Fluoxetine with only three fluorine atoms ($C_{17}H_{18}F_3NO$, m/C ≈ 18) fall below the applied cutoff. Depending on the underlying NTS question, this cutoff can be adjusted accordingly. Attempting to remove further features, an MD/C cutoff of < +0.003 was set, although as seen in Fig. **3a** the m/C dimension was much more effective for data reduction. The MD/C-m/C approach was more efficient to reduce features compared to the MD, as shown in Fig. **3b** and **3d**. When applying a MD range from -0.25 to +0.1 Da, which would include 92% of the PFAS in the PFASOECDNA list (EPA dashboard), 17% of the features remained, while the combined m/C and MD/C cutoffs led to only 7.4% of remaining features. It is very important to note here that the number of features that strongly exceed a MD of +0.5 is not negligible, since a conventional calculation of the MD would result in a negative MD (e.g., -0.2 Da for a saturated hydrocarbon with 60 carbon atoms ($H(CH_2)_{60}H$), whereas the true MD would be +0.8 Da). As can be seen from the carbon number, a considerable number of features has more than 60 carbon atoms (up to 80 carbons) which are in a PFAS typical MD-range (Fig. **3b**). Therefore, setting an appropriate m/C cutoff is highly recommended, since these features are easily removed by this additional criterion. Eventually, when combining both m/C and MD/C cutoffs, only 949 features (7.4%) remain in all four soils together. This is an appropriate number of features for further PFAS specific calculations such as KMD analysis, DFs, fragment mass differences and suspect screening. It should be noted in particular that due to the removal of ~90% of the initial features, the false-positive rate decreases drastically (especially with large lists), and allows adjustment of selected tolerances with smaller effect on false positives.

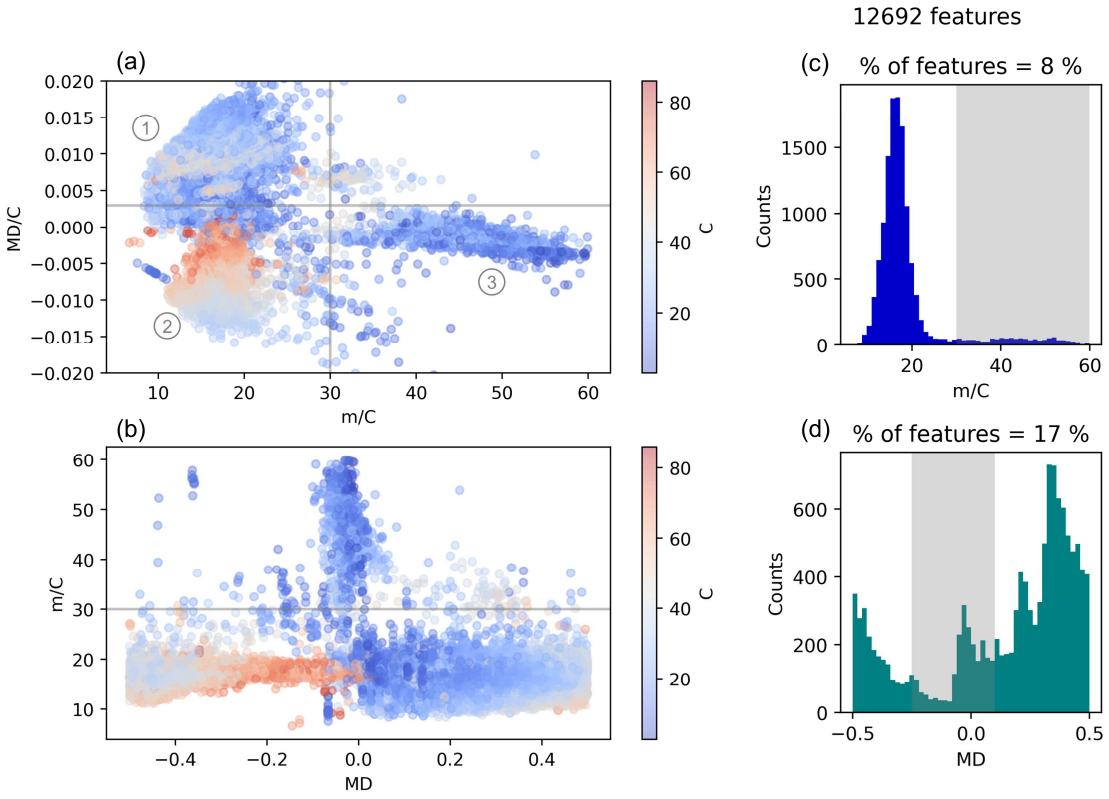


Fig. 3: Data reduction by the MD/C-m/C approach compared to the MD. (a) MD/C-m/C plot for all detected features (12692) in the four soil extracts and (b) m/C vs. MD. The colorbars correspond to the calculated carbon number. In the MD/C-m/C plot, potential PFAS (3) are clearly separated from hydrocarbons (1) and hydrocarbons with many carbon atoms that exceed a MD of +0.5 and are therefore flipped to a similar MD region as the PFAS but are easily separated by m/C. The number of features is reduced to 8% by the m/C dimension when cutting at m/C > 30 and to 7.4% when including a threshold of MD/C < 0.003 (grey lines in subplot a). (c) Histogram of m/C and of the MD (d), showing that the m/C works more efficiently than the MD (17% of the features remain when cutting at -0.25 < MD > 0.1 which includes 92% of the PFASOECDNA list [27]). Many features strongly exceeding a MD of +0.5 would be wrongly prioritized. Note how the m/C dimension allows a much clearer cutoff from hydrocarbon-based features compared to the MD.

KMD analysis, fragment differences, DFs and suspect screening

For further prioritization and tentative identification, repeating units representative for PFAS such as $(CF_2)_n$ and CF_2O were applied to detect HS (mass tolerance was set at ± 2 mDa, with at least 3 homologues). Without any m/C cutoff, in soil M1, 74 $(CF_2)_n$-based HS were detected, likely including numerous false-positives (Fig. 4a) evidenced by a random RT pattern (no RT-shift in linked KMD m/z vs. RT plot). The KMD analysis in PFΔ*Screen* is performed without checking the systematic RT-shift, but the interactive KMD plot (HTML) allow a fast verification of RT-shifts. Each HS can be highlighted individually by clicking on it, and the respective m/z vs. RT correlation is visualized (Fig. S5). Obviously, many hydrocarbon

features were detected in the soil extract that are mimicking CF_2-repeating units, which is a common issue of complex matrices [31,37]. These compounds have a higher CF_2-based KMD (e.g., 0.2 to 0.5, or lower if their MD strongly exceeds +0.5 Da) compared to that of PFAS (Fig. 4a). If the combined MD/C-m/C cutoff is applied, the number of detected HS in soil M1 is reduced to 26 (~65% data reduction, see Fig. 4b) which confirms the utility of this approach.

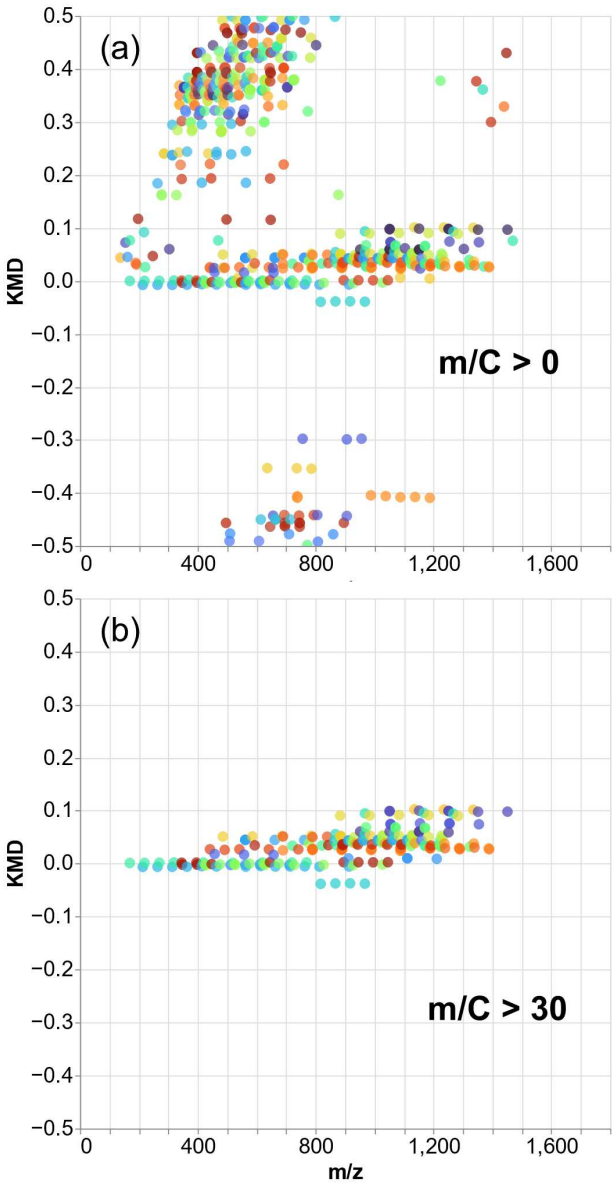


Fig. 4: True and false-positive CF_2-based HS in soil M1 (a) without (m/C > 0) and (b) with m/C-cutoff (m/C > 30). An MS^1 noise threshold of 1000 counts was used for feature detection, and the KMD mass tolerance was set to ± 1 mDa with a minimum of three homologues. Even with the low mass tolerance of ± 1 mDa many hydrocarbon matrix components are mimicking the CF_2-repeating unit (see also Fig S5). Note: Multiple $(CF_2)_n$ differences within the KMD tolerance are also assigned to the respective HS, therefore each datapoint has at least two HS partners.

For detection of fragment mass differences and DFs in the MS^2 data, preliminary ΔCF_2, ΔC_2F_4, ΔHF and the list of DFs were used (later specific mass differences were searched). This resulted in the detection of 30 MS^2 spectra that contained the specified mass differences, and 47 spectra with DF hits out of a total number of 373 unique MS^2 spectra at a mass tolerance set to ± 2 mDa and an MS^2 intensity threshold of 2000 counts in the M1 soil extract (first iteration).

In the suspect screening process, the hits by accurate mass (tolerance of 4 mDa) were reduced from 217 to 176 by the MD/C-m/C cutoff in soil M1.

Manual identification process with the PFΔ*Screen* results table

The verification and (partially manual) identification process of prioritized features from the PFΔ*Screen* results table (Excel) was performed by sorting the table according to decreasing intensity, after removing features based on defined MD/C-m/C cutoffs. For soil M1, this resulted in a feature list with 305 potential compounds. Note that some features appear multiple times in the list due to structural isomerism, resulting in multiple features at multiple distinct RTs depending on the degree of separation and the peak finding algorithm. Each feature was verified manually for occurrence in the extraction blank and reasonable peak shape (until <1% of the most abundant feature). Although a blank correction was performed, typical contaminations from the LC system with long tailing peaks can be integrated multiple times at different RTs. Therefore, they are not always correctly removed depending on the specified parameters. By using the RawDataVisualization tool of PFΔ*Screen,* EICs of every m/z belonging to one HS (using the integrated HS extrapolator) can be verified for RT-shift and peak shape, eventually resulting in identification of homologues with very low abundances that were missed in the feature finding process due to the MS^1 intensity threshold. The chemical formulas from suspect hits were used to check for reasonable isotope patterns with

the RawDataVisualization of PFΔ*Screen.* SMILES codes were used to verify at least one candidate per HS by an MS^2 spectrum.

In total, nine PFAS classes could be identified via PFΔ*Screen* in the four soils that exhibited at least one suspect hit per HS or compound (Fig. 5). Perfluoroalkyl carboxylic acids (PFCAs, C_4 – C_{20}), fluorotelomer alkyl phosphate diesters (diPAPs, 4:2/6:2 – 12:2/12:2), n:3 fluorotelomer carboxylic acids (FTCAs, 5:3-13:3), fluorotelomer sulfonic acids (FTSAs, 6:2 – 16:2), perfluorosulfonic acids (PFSAs, C_4 – C_{10}), perfluorooctane sulfonamide (PFOSA), N-ethylperfluoro-1-octanesulfonamidoacetic acid (N-EtFOSAA), and N-ethyl perfluorooctane sulfonamide ethanol–based phosphate diester (diSAmPAP) were identified in all four soils. Different chain-length distributions and abundances were observed (Fig. 5). diPAPs were detected as complex mixtures of several structural isomers depending on their chain-length (e.g., 6:2/10:2 and 8:2/8:2, shown by MS/MS). Additionally, their EICs showed peaks at much later RTs corresponding to in-source fragments of triPAPs (Fig. S7). While all telomer-based PFAS were detected as linear chains, the PASF-based PFAS (PFSAs, N-EtFOSAA, PFOSA, and diSAmPAP) showed typical chromatographic peak shapes of mixtures of branched and linear isomers [55]. In these cases, the dominance of a C_8-based chemistry can be observed (see PFSAs in Fig. 5).

All four soils had a similar contamination pattern. However, for soils M1 and M2 (Mannheim region) another very abundant precursor class, namely FTMAPs, were detected (including isomeric profiles ranging 6:2/6:2 – 10:2/12:2), as well the previously identified TPs FTMAP-sulfoxides [31].

The 6:2 fluorotelomer mercapto alkyl phosphate esters (6:2/6:2 FTMAP) could be confirmed with an in-house synthesized reference standard, leading to identification levels of 1 for 6:2 FTMAP and 2a for the further homologues due to clear MS/MS evidence [56]. In general, all identified PFAS are in good agreement with previous studies including

biotransformation that characterized other soil samples from both Rastatt and Mannheim [27,32,31,38,54].

The PFΔ*Screen* results table also revealed several unknown HS that were detected but did not have an accurate mass match with the suspect list. Their identification with the help of the EIC correlator of PFΔ*Screen* is discussed in the following.

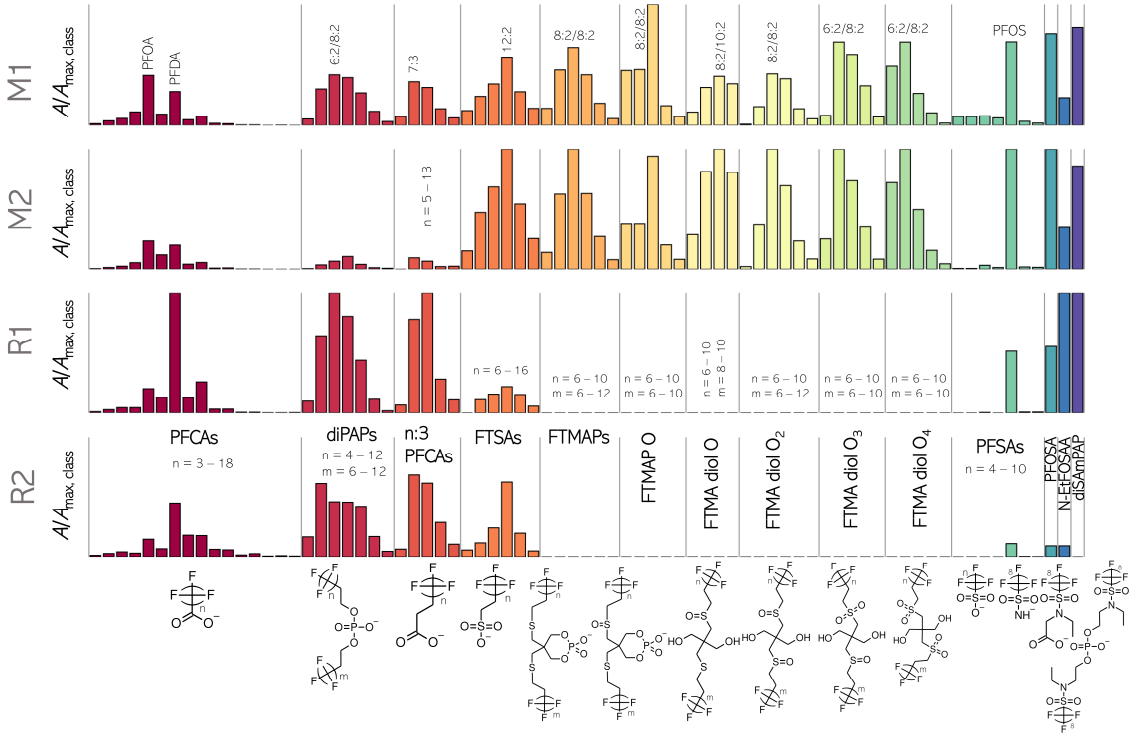


Fig. 5: Qualitative summary of identified PFAS in the four soils (M1, M2, R1, and R2). Each class (e.g., PFCAs, diPAPs) is normalized to the peak area of the most abundant homologue within all four samples. Further abbreviations: FTMAP O: FTMAP-sulfoxide, FTMA diol O: FTMA-diol-sulfoxide, FTMA diol O_2: FTMA-diol-disulfoxid or -sulfone, FTMA diol O_3: FTMA-diol-sulfoxide-sulfone, FTMA diol O_4: FTMA-diol-disulfone, Note that depending on chain length and sulfur oxidation degree, diPAPs, FTMAPs and FTMA-diols were detected as complex mixture of structural and positional isomers (e.g., 6:2/10:2 ≠ 8:2/8:2, or disulfoxide ≠ sulfone). Very small abundant identifications and triPAPs are not shown in Fig. 5.

EIC correlator: Coelution correlation analysis for identification of unknowns

After identification of the PFA*Screen* results, there were several C_2F_4-based HS left without any hit in the suspect list. When looking at several MS^1 spectra of different homologues, many coeluting ions were observed, often characterized by HF losses and other mass differences (Fig. S8). This is an indication of in-source fragmentation of these classes [57,58]. To be able to efficiently group corresponding in-source fragments and potential adduct ions together, the

EIC correlator from the raw data visualization tools of PFΔ*Screen* was used to correlate the EICs of suspected features (from a given HS) with the EICs of all detected features that coelute within a given RT-range of ± 25 seconds. Strong correlation of EICs can be used to detect related ions and allows their isolation from other ions in consecutive MS^1 spectra without knowing their mass differences [59-62]. This is exemplified on the unknown m/z 966.9944 which is a member of a suspected HS. When correlating the EIC of m/z = 966.9944 with the EICs of all coeluting features within a RT range of 50 seconds, 12 out of 368 EICs correlated with an $R^2 > 0.96$ at an extraction width of 5 mDa (see Fig **S9** for more details). The result is an MS^1 spectrum that only contains co-eluting ions (correlation spectrum) of several in-source fragments and adducts (Fig. **6**). Since well-known mass differences such as ΔC_2F_4 and ΔHF were found in this MS^1 spectrum, a telomer-based PFAS with potentially two telomer-chains (e.g., 6:2/8:2) was suspected [31]. When looking at the mass differences of detected coeluting ions, $[M+Cl]^-$, $[M+Br]^-$, $[M+Ac]^-$ adducts and several other in-source fragments could be observed. The detection of $[M+Cl]^-$ and $[M+Br]^-$ ions were of great importance since they allowed the determination of [M] rather easily which then also allowed the identification of other adducts and the molecular formula. The m/z = 966.9944 (in-source fragment) corresponds to a FTMAP related substance, which was tagged FTMA-diol-sulfone-sulfoxide or FTMA-diol-O_3 (see Fig. **5** and **6**). With this correlation technique, several tens of unknown HS could be grouped into four novel FTMAP related compound classes (Fig. **6**). They were identified with one oxygen (sulfoxide) and up to 4 oxygens (disulfone) and to the best of our knowledge not reported in literature before. They could be microbial or photochemical FTMAP TPs and close the unknown gap in a previous FTMAP-related transformation study [63], or they could be used intentionally or as side-products in PFAS-coated papers that contaminate these soils. These kind of correlation spectra made identification possible since the MS^2 spectra of the adducts ($[M-H]^-$ ions of the FTMA-diols were not detected at all which makes sense with ESI) barely formed useful fragments except for Br^- which made them hard

to interpret. The use of in-source fragments for identification has the advantage that isotope patterns are available for all ions (features), which is often not the case in MS^2 spectra depending on the isolation width of the precursor ion. All these FTMAP-related substances form multiple in-source fragments (and adducts), all could be confirmed with rather high confidence (identification level of 2b). They all could be grouped by C_2F_4- and O-based KMD (for O-KMD see Fig. **S10**) with systematic RT-shifts, besides eluting at higher RT than FTMAPs due to their lower polarity attributed to the loss of the phosphoric acid group.

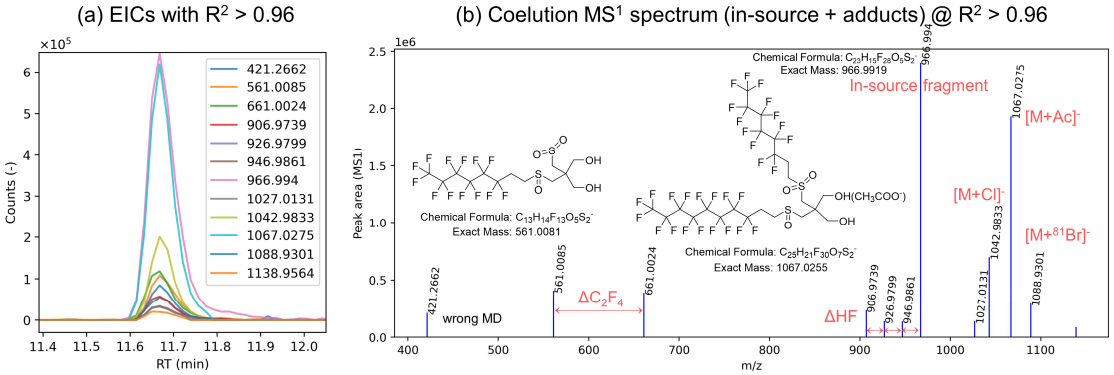


Fig. 6: Detection of coeluting in-source fragments and adducts via the EIC correlator of PFΔ*Screen* for the identification of 6:2/8:2 FTMA diol sulfoxide sulfone. The EIC of the unknown insource fragment m/z = 966.9944 (which was detected as one member of a HS via KMD) was correlated with all EICs eluting at its RT ± 25 s resulting in non-targeted detection of related ions. In total, 4 HS corresponding to 21 novel FTMAP TPs were identified via the use of this tool (see Fig. **5**). A RT-shift with increasing oxidation degree (1 O up to 4 O) was observed due to increasing polarity. Note that the EICs of [M+^{37}Cl] and [M+^{79}Cl] are also in the raw MS^1 spectra, however they were combined into one feature by feature finding algorithm of pyOpenMS (in case of Br a wrong isotope grouping occurred) and therefore not detectable by the correlation analysis.

Conclusions

PFΔ*Screen* can efficiently be used for prioritizing features in both LC- and GC- HRMS raw data in all kinds of samples independently of the vendor of the mass spectrometer used. Especially, the MD/C-m/C approach is a powerful tool to drastically decrease the number of features and thus reduce false-positive assignments, overcoming a common issue during NTS. Due to the short computational time of PFΔ*Screen* (less than one minute for 4000 spectra), input parameters can be conveniently adjusted depending on the tested sample, instrument used and end-user needs. Since the number of unknown PFAS in complex environmental and technical samples is still unknown, NTS approaches that combine several data reduction

techniques for an efficient workflow are of importance to comprehensively elucidate the identity occurrence and fate of organic pollutants such as PFAS.

Associated Content

Supporting information: ESI as PDF.

Author contributions

JZ conceptualized the structure of PFΔ*Screen*, wrote most of the Python source code and wrote the first draft of the manuscript. BB was part of writing and designing PFΔ*Screen* processes and reviewed the manuscript. JFP performed the soil extractions and reviewed the manuscript. CZ supervised the study and reviewed the manuscript.

Author information

Corresponding Author:

*Jonathan Zweigle – jonathan.zweigle@uni-tuebingen.de

*Christian Zwiener – christian.zwiener@uni-tuebingen.de

Notes

The authors declare no competing financial interest.

Code availability and license

The Python source code of PFΔ*Screen* is available on GitHub (https://github.com/JonZwe/PFAScreen) together with example files. It is published under the LGPL-2.1 license.

Acknowledgements

The authors would like to thank Axel Walter for his GitHub repository *ion-chromatogram-extractor* which inspired some functions related to building EICs in PFΔ*Screen*, Klaus Röhler for sampling the composite soil samples and the DBU (Deutsche Bundesstiftung Umwelt) for the scholarship of JZ.

References

1. Evich MG, Davis MJB, McCord JP, Acrey B, Awkerman JA, Knappe DRU, Lindstrom AB, Speth TF, Tebes-Stevens C, Strynar MJ, Wang Z, Weber EJ, Henderson WM, Washington JW (2022) Per- and polyfluoroalkyl substances in the environment. Science 375 (6580):eabg9065. doi:10.1126/science.abg9065
2. Lindstrom AB, Strynar MJ, Libelo EL (2011) Polyfluorinated compounds: past, present, and future. Environ Sci Technol 45 (19):7954-7961. doi:10.1021/es2011622
3. Ng C, Cousins IT, DeWitt JC, Glüge J, Goldenman G, Herzke D, Lohmann R, Miller M, Patton S, Scheringer M, Trier X, Wang Z (2021) Addressing Urgent Questions for PFAS in the 21st Century. Environ Sci Technol. doi:10.1021/acs.est.1c03386
4. Glüge J, Scheringer M, Cousins IT, DeWitt JC, Goldenman G, Herzke D, Lohmann R, Ng CA, Trier X, Wang Z (2020) An overview of the uses of per- and polyfluoroalkyl substances (PFAS). Environ Sci Process Impacts 22 (12):2345-2373. doi:10.1039/d0em00291g
5. Cousins IT, DeWitt JC, Glüge J, Goldenman G, Herzke D, Lohmann R, Ng CA, Scheringer M, Wang Z (2020) The high persistence of PFAS is sufficient for their management as a chemical class. Environ Sci Process Impacts 22 (12):2307-2312. doi:10.1039/d0em00355g
6. Wang Z, Cousins IT, Scheringer M, Buck RC, Hungerbuhler K (2014) Global emission inventories for C4-C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030, Part I: production and emissions from quantifiable sources. Environ Int 70:62-75. doi:10.1016/j.envint.2014.04.013
7. Wang Z, Cousins IT, Scheringer M, Buck RC, Hungerbuhler K (2014) Global emission inventories for C4-C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030, part II: the remaining pieces of the puzzle. Environ Int 69:166-176. doi:10.1016/j.envint.2014.04.006
8. Cousins IT, Johansson JH, Salter ME, Sha B, Scheringer M (2022) Outside the Safe Operating Space of a New Planetary Boundary for Per- and Polyfluoroalkyl Substances (PFAS). Environ Sci Technol 56 (16):11172-11179. doi:10.1021/acs.est.2c02765
9. Wang Z, Buser AM, Cousins IT, Demattio S, Drost W, Johansson O, Ohno K, Patlewicz G, Richard AM, Walker GW, White GS, Leinala E (2021) A New OECD Definition for Per- and Polyfluoroalkyl Substances. Environ Sci Technol 55 (23):15575-15578. doi:10.1021/acs.est.1c06896
10. Schymanski E, Zhang J, Thiessen PA, Chirsir P, Kondic T, Bolton EE (2023) Per- and polyfluoroalkyl substances (PFAS) in PubChem: 7 million and growing.
11. Stockholm Convention (2022) The new POPs under the Stockholm Convention. http://www.pops.int/TheConvention/ThePOPs/TheNewPOPs/tabid/2511/Default.aspx. Accessed 21.03.2023
12. Kwiatkowski CF, Andrews DQ, Birnbaum LS, Bruton TA, DeWitt JC, Knappe DRU, Maffini MV, Miller MF, Pelch KE, Reade A, Soehl A, Trier X, Venier M, Wagner CC, Wang Z, Blum A (2020) Scientific Basis for Managing PFAS as a Chemical Class. Environ Sci Technol Lett 7 (8):532-543. doi:10.1021/acs.estlett.0c00255
13. ECHA (2023) ECHA publishes PFAS restriction proposal. https://echa.europa.eu/de/-/echa-publishes-pfas-restriction-proposal. Accessed 21.09.2023
14. Aro R, Carlsson P, Vogelsang C, Karrman A, Yeung LW (2021) Fluorine mass balance analysis of selected environmental samples from Norway. Chemosphere 283:131200. doi:10.1016/j.chemosphere.2021.131200
15. Aro R, Eriksson U, Kärrman A, Chen F, Wang T, Yeung LWY (2021) Fluorine Mass Balance Analysis of Effluent and Sludge from Nordic Countries. ACS ES&T Water 1 (9):2087-2096. doi:10.1021/acsestwater.1c00168
16. Simon F, Gehrenkemper L, Becher S, Dierkes G, Langhammer N, Cossmer A, von der Au M, Gockener B, Fliedner A, Rudel H, Koschorreck J, Meermann B (2023) Quantification and characterization of PFASs in suspended particulate matter (SPM) of German rivers using EOF, dTOPA, (non-)target HRMS. Sci Total Environ 885:163753. doi:10.1016/j.scitotenv.2023.163753
17. Koch A, Aro R, Wang T, Yeung LWY (2020) Towards a comprehensive analytical workflow for the chemical characterisation of organofluorine in consumer products and environmental samples. Trac-Trend Anal Chem 123:115423. doi:10.1016/j.trac.2019.02.024

18. Aro R, Eriksson U, Karrman A, Yeung LWY (2021) Organofluorine Mass Balance Analysis of Whole Blood Samples in Relation to Gender and Age. Environ Sci Technol 55 (19):13142-13151. doi:10.1021/acs.est.1c04031
19. Ruan T, Jiang G (2017) Analytical methodology for identification of novel per- and polyfluoroalkyl substances in the environment. TrAC Trends in Analytical Chemistry 95:122-131. doi:10.1016/j.trac.2017.07.024
20. Jia S, Marques Dos Santos M, Li C, Snyder SA (2022) Recent advances in mass spectrometry analytical techniques for per- and polyfluoroalkyl substances (PFAS). Anal Bioanal Chem 414 (9):2795-2807. doi:10.1007/s00216-022-03905-y
21. Strynar M, McCord J, Newton S, Washington J, Barzen-Hanson K, Trier X, Liu Y, Dimzon IK, Bugsel B, Zwiener C, Munoz G (2023) Practical application guide for the discovery of novel PFAS in environmental samples using high resolution mass spectrometry. J Expo Sci Environ Epidemiol. doi:10.1038/s41370-023-00578-2
22. Joerss H, Menger F (2023) The complex 'PFAS world' - how recent discoveries and novel screening tools reinforce existing concerns. Current Opinion in Green and Sustainable Chemistry. doi:10.1016/j.cogsc.2023.100775
23. Zweigle J, Bugsel B, Röhler K, Haluska AA, Zwiener C (2023) PFAS-Contaminated Soil Site in Germany: Nontarget Screening before and after Direct TOP Assay by Kendrick Mass Defect and FindPFΔS. Environ Sci Technol 57 (16):6647-6655. doi:10.1021/acs.est.2c07969
24. Liu Y, D'Agostino LA, Qu G, Jiang G, Martin JW (2019) High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly- and per-fluoroalkyl substances (PFASs) in environmental and human samples. TrAC Trends in Analytical Chemistry 121. doi:10.1016/j.trac.2019.02.021
25. Hulleman T, Turkina V, O'Brien JW, Chojnacka A, Thomas KV, Samanipour S (2023) Critical Assessment of the Chemical Space Covered by LC-HRMS Non-Targeted Analysis. Environ Sci Technol. doi:10.1021/acs.est.3c03606
26. Zweigle J, Bugsel B, Zwiener C (2023) Efficient PFAS prioritization in non-target HRMS data: systematic evaluation of the novel MD/C-m/C approach. Analytical and Bioanalytical Chemistry. doi:10.1007/s00216-023-04601-1
27. Bugsel B, Zwiener C (2020) LC-MS screening of poly- and perfluoroalkyl substances in contaminated soil by Kendrick mass analysis. Anal Bioanal Chem 412 (20):4797-4805. doi:10.1007/s00216-019-02358-0
28. Koelmel JP, Paige MK, Aristizabal-Henao JJ, Robey NM, Nason SL, Stelben PJ, Li Y, Kroeger NM, Napolitano MP, Savvaides T, Vasiliou V, Rostkowski P, Garrett TJ, Lin E, Deigl C, Jobst K, Townsend TG, Godri Pollitt KJ, Bowden JA (2020) Toward Comprehensive Per- and Polyfluoroalkyl Substances Annotation Using FluoroMatch Software and Intelligent High-Resolution Tandem Mass Spectrometry Acquisition. Analytical Chemistry 92 (16):11186-11194. doi:10.1021/acs.analchem.0c01591
29. Dickman RA, Aga DS (2022) Efficient workflow for suspect screening analysis to characterize novel and legacy per- and polyfluoroalkyl substances (PFAS) in biosolids. Anal Bioanal Chem. doi:10.1007/s00216-022-04088-2
30. Kaufmann A, Butcher P, Maden K, Walker S, Widmer M (2022) Simplifying Nontargeted Analysis of PFAS in Complex Food Matrices. J AOAC Int. doi:10.1093/jaoacint/qsac071
31. Zweigle J, Bugsel B, Zwiener C (2022) FindPFΔS: Non-Target Screening for PFAS - Comprehensive Data Mining for MS2 Fragment Mass Differences. Anal Chem 94 (30):10788-10796. doi:10.1021/acs.analchem.2c01521
32. Bugsel B, Bauer R, Herrmann F, Maier ME, Zwiener C (2022) LC-HRMS screening of per- and polyfluorinated alkyl substances (PFAS) in impregnated paper samples and contaminated soils. Anal Bioanal Chem 414 (3):1217-1225. doi:10.1007/s00216-021-03463-9
33. Munoz G, Michaud AM, Liu M, Vo Duy S, Montenach D, Resseguier C, Watteau F, Sappin-Didier V, Feder F, Morvan T, Houot S, Desrosiers M, Liu J, Sauve S (2022) Target and Nontarget Screening of PFAS in Biosolids, Composts, and Other Organic Waste Products for Land Application in France. Environ Sci Technol 56 (10):6056-6068. doi:10.1021/acs.est.1c03697
34. Liu L, Lu M, Cheng X, Yu G, Huang J (2022) Suspect screening and nontargeted analysis of per- and polyfluoroalkyl substances in representative fluorocarbon surfactants, aqueous film-forming

foams, and impacted water in China. Environment International 167. doi:10.1016/j.envint.2022.107398
35. Ng K, Alygizakis N, Androulakakis A, Galani A, Aalizadeh R, Thomaidis NS, Slobodnik J (2022) Target and suspect screening of 4777 per- and polyfluoroalkyl substances (PFAS) in river water, wastewater, groundwater and biota samples in the Danube River Basin. J Hazard Mater 436:129276. doi:10.1016/j.jhazmat.2022.129276
36. Grulke CM, Williams AJ, Thillanadarajah I, Richard AM (2019) EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. Comput Toxicol 12. doi:10.1016/j.comtox.2019.100096
37. Young RB, Pica NE, Sharifan H, Chen H, Roth HK, Blakney GT, Borch T, Higgins CP, Kornuc JJ, McKenna AM, Blotevogel J (2022) PFAS Analysis with Ultrahigh Resolution 21T FT-ICR MS: Suspect and Nontargeted Screening with Unrivaled Mass Resolving Power and Accuracy. Environ Sci Technol 56 (4):2455-2465. doi:10.1021/acs.est.1c08143
38. Nürenberg G, Nödler K, T LF, Schäfer C, Huber K, Scheurer M (2018) Nachweis von polyfluorierten Alkylphosphatestern (PAP) und Perfluoroktansulfonamidoethanol-basierten Phosphatestern (SAmPAP) in Böden. Mitt Umweltchem Ökotox
39. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, 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 EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30 (10):918-920. doi:10.1038/nbt.2377
40. Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Rompp A, Neumann S, Pizarro AD, Montecchi-Palazzi L, Tasman N, Coleman M, Reisinger F, Souda P, Hermjakob H, Binz PA, Deutsch EW (2011) mzML--a community standard for mass spectrometry data. Mol Cell Proteomics 10 (1):R110 000133. doi:10.1074/mcp.R110.000133
41. Röst HL, Schmitt U, Aebersold R, Malmstrom L (2014) pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics 14 (1):74-77. doi:10.1002/pmic.201300246
42. Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich HC, Gutenbrunner P, Kenar E, Liang X, Nahnsen S, Nilse L, Pfeuffer J, Rosenberger G, Rurik M, Schmitt U, Veit J, Walzer M, Wojnar D, Wolski WE, Schilling O, Choudhary JS, Malmstrom L, Aebersold R, Reinert K, Kohlbacher O (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods 13 (9):741-748. doi:10.1038/nmeth.3959
43. Sturm M, Bertsch A, Gropl C, Hildebrandt A, Hussong R, Lange E, Pfeifer N, Schulz-Trieglaff O, Zerck A, Reinert K, Kohlbacher O (2008) OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics 9:163. doi:10.1186/1471-2105-9-163
44. Sachsenberg T, Pfeuffer J, Bielow C, Wein S, Jeong K, Netz E, Walter A, Alka O, Nilse L, Colaianni P, McCloskey D, Kim J, Rosenberger G, Bichmann L, Walzer M, Veit J, Boudaud B, Bernt M, Patikas N, Pilz M, Startek MP, Kutuzova S, Heumos L, Charkow J, Sing J, Feroz A, Siraj A, Weisser H, Dijkstra T, Perez-Riverol Y, Röst H, Kohlbacher O (2023) OpenMS 3 expands the frontiers of open-source computational mass spectrometry. Preprint. doi:10.21203/rs.3.rs-3286368/v1
45. Pfeuffer J, Sachsenberg T, Alka O, Walzer M, Fillbrunn A, Nilse L, Schilling O, Reinert K, Kohlbacher O (2017) OpenMS - A platform for reproducible analysis of mass spectrometry data. J Biotechnol 261:142-148. doi:10.1016/j.jbiotec.2017.05.016
46. Kenar E, Franken H, Forcisi S, Wormann K, Haring HU, Lehmann R, Schmitt-Kopplin P, Zell A, Kohlbacher O (2014) Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data. Mol Cell Proteomics 13 (1):348-359. doi:10.1074/mcp.M113.031278
47. Helmus R, Ter Laak TL, van Wezel AP, de Voogt P, Schymanski EL (2021) patRoon: open source software platform for environmental mass spectrometry based non-target screening. J Cheminform 13 (1):1. doi:10.1186/s13321-020-00477-w
48. Kontou EE, Walter A, Alka O, Pfeuffer J, Sachsenberg T, Mohite OS, Nuhamunada M, Kohlbacher O, Weber T (2023) UmetaFlow: an untargeted metabolomics workflow for high-throughput data processing and analysis. J Cheminform 15 (1):52. doi:10.1186/s13321-023-00724-w

49. Goloborodko AA, Levitsky LI, Ivanov MV, Gorshkov MV (2013) Pyteomics--a Python framework for exploratory data analysis and rapid software prototyping in proteomics. J Am Soc Mass Spectrom 24 (2):301-304. doi:10.1007/s13361-012-0516-6
50. Levitsky LI, Klein JA, Ivanov MV, Gorshkov MV (2019) Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. J Proteome Res 18 (2):709-714. doi:10.1021/acs.jproteome.8b00717
51. Koelmel JP, Stelben P, McDonough CA, Dukes DA, Aristizabal-Henao JJ, Nason SL, Li Y, Sternberg S, Lin E, Beckmann M, Williams AJ, Draper J, Finch JP, Munk JK, Deigl C, Rennie EE, Bowden JA, Godri Pollitt KJ (2022) FluoroMatch 2.0-making automated and comprehensive non-targeted PFAS annotation a reality. Anal Bioanal Chem 414 (3):1201-1215. doi:10.1007/s00216-021-03392-7
52. Barzen-Hanson KA, Roberts SC, Choyke S, Oetjen K, McAlees A, Riddell N, McCrindle R, Ferguson PL, Higgins CP, Field JA (2017) Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater. Environ Sci Technol 51 (4):2047-2057. doi:10.1021/acs.est.6b05843
53. Xiao F, Golovko SA, Golovko MY (2017) Identification of novel non-ionic, cationic, zwitterionic, and anionic polyfluoroalkyl substances using UPLC-TOF-MS(E) high-resolution parent ion search. Anal Chim Acta 988:41-49. doi:10.1016/j.aca.2017.08.016
54. Röhler K, Susset B, Grathwohl P (2023) Production of perfluoroalkyl acids (PFAAs) from precursors in contaminated agricultural soils: Batch and leaching experiments. Sci Total Environ 902:166555. doi:10.1016/j.scitotenv.2023.166555
55. Londhe K, Lee C-S, McDonough CA, Venkatesan AK (2022) The Need for Testing Isomer Profiles of Perfluoroalkyl Substances to Evaluate Treatment Processes. Environmental Science & Technology. doi:10.1021/acs.est.2c05518
56. Charbonnet JA, McDonough CA, Xiao F, Schwichtenberg T, Cao D, Kaserzon S, Thomas KV, Dewapriya P, Place BJ, Schymanski EL, Field JA, Helbling DE, Higgins CP (2022) Communicating Confidence of Per- and Polyfluoroalkyl Substance Identification via High-Resolution Mass Spectrometry. Environ Sci Technol Lett 9 (6):473-481. doi:10.1021/acs.estlett.2c00206
57. Berger U, Langlois I, Oehme M, Kallenborn R (2004) Comparison of three types of mass spectrometers for HPLC/MS analysis of perfluoroalkylated substances and fluorotelomer alcohols. Eur J Mass Spectrom (Chichester) 10 (5):579-588. doi:10.1255/ejms.679
58. Trier X, Granby K, Christensen JH (2011) Tools to discover anionic and nonionic polyfluorinated alkyl surfactants by liquid chromatography electrospray ionisation mass spectrometry. Journal of Chromatography A 1218 (40):7094-7104. doi:10.1016/j.chroma.2011.07.057
59. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84 (1):283-289. doi:10.1021/ac202450g
60. Godzien J, Armitage EG, Angulo S, Martinez-Alcazar MP, Alonso-Herranz V, Otero A, Lopez-Gonzalvez A, Barbas C (2015) In-source fragmentation and correlation analysis as tools for metabolite identification exemplified with CE-TOF untargeted metabolomics. Electrophoresis 36 (18):2188-2195. doi:10.1002/elps.201500016
61. Seitzer PM, Searle BC (2019) Incorporating In-Source Fragment Information Improves Metabolite Identification Accuracy in Untargeted LC-MS Data Sets. J Proteome Res 18 (2):791-796. doi:10.1021/acs.jproteome.8b00601
62. Tada I, Chaleckis R, Tsugawa H, Meister I, Zhang P, Lazarinis N, Dahlen B, Wheelock CE, Arita M (2020) Correlation-Based Deconvolution (CorrDec) To Generate High-Quality MS2 Spectra from Data-Independent Acquisition in Multisample Studies. Anal Chem 92 (16):11310-11317. doi:10.1021/acs.analchem.0c01980
63. Bugsel B, Schussler M, Zweigle J, Schmitt M, Zwiener C (2023) Photocatalytical transformation of fluorotelomer- and perfluorosulfonamide-based PFAS on mineral surfaces and soils in aqueous suspensions. Sci Total Environ 894:164907. doi:10.1016/j.scitotenv.2023.164907