Science of the Total Environment 908 (2024) 167967

Contents lists available at ScienceDirect

Science of the Total Environment

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

Assessment for the data processing performance of non-target screening analysis based on high-resolution mass spectrometry

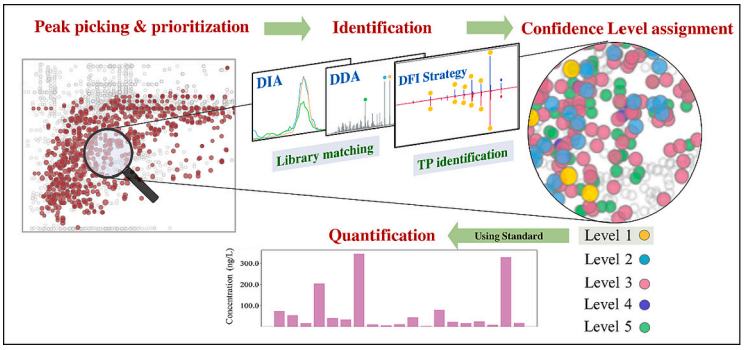
He Liu [1], Rui Wang [1], Bo Zhao [*], Danping Xie [*]

South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, PR China
State Environmental Protection Key Laboratory of Water Environmental Simulation and Pollution Control, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, PR China

HIGHLIGHTS

- Performance evaluation of a data processing workflow for non-target screening (NTS)
- Accurate structure elucidation for DIA and DDA data in NTS
- Insights into the application of NTS workflow to real water samples

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Thomas Kevin V

Keywords:
NTS
PyHRMS
Contaminants of emerging concern
DDA
DIA

ABSTRACT

Non-target screening (NTS) based on high-resolution mass spectrometry (HRMS) is considered one of the most comprehensive approaches for the characterization of contaminants of emerging concern (CECs) in a complex sample. This study evaluated the performance of NTS in aquatic environments (including peak picking, database matching, product identification, semi-quantification, etc.) based on a self-developed data processing method using 38 glucocorticoids as testing compounds. Data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes were used for obtaining the MS^2 information for in-house or online database matching. Results indicate that DDA and DIA mode have their own advantages and can complement each other. The quantification method based on LC-HRMS has shown the potential to provide a fast and acceptable result for testing compounds. Finally, a matrix spike analysis was carried out on 66 CECs across different usage categories in wastewater, surface water, and seawater matrix samples, together with a case study performed for characterizing the whole contaminants in a Pearl River sample, to better illustrate the application potential of NTS workflow and the credibility of NTS outcomes. This study provides a foundation for novel applications of HRMS data by NTS workflow to identify and quantify CECs in complex systems.

* Corresponding authors at: South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, PR China.
E-mail addresses: zhaobo@scies.org (B. Zhao), xiedanping@scies.org (D. Xie).
[1] H.L. and R.W. contributed equally to this paper.

https://doi.org/10.1016/j.scitotenv.2023.167967
Received 10 February 2023; Received in revised form 17 October 2023; Accepted 18 October 2023
Available online 20 October 2023

1. Introduction

The increasing worldwide release of anthropogenic chemicals into the aquatic environment has caused serious contamination of human freshwater resources (Petrie et al., 2015; Tian et al., 2020). Some of these chemicals have raised considerable toxicological concerns even present at very low concentrations (Tian et al., 2021). The structural elucidation of these contaminants of emerging concern (CECs), particularly when present in a complex environmental matrix, is one of the key challenges facing environmental analysts (Fisher et al., 2022). Non-target screening (NTS) approach based on liquid chromatography-high resolution mass spectrometry (LC-HRMS) has been widely applied to characterize the organic contaminants in the aquatic environment (Fisher et al., 2022; Hollender et al., 2017a). HRMS is capable of providing the accurate mass of thousands of substances with a fast scanning rate, which facilitates compound identification without authentic standards (Schymanski et al., 2014b). In addition, the high throughput detecting ability of LC-HRMS can acquire comprehensive knowledge of complex environmental samples, making it practicable to efficiently recognize all existing chemicals and digitally record their information for retrospective exploration (Alygizakis et al., 2018; Günthardt et al., 2021). The NTS approach, as a complementary tool to target analysis, can help alleviate the monitoring burden caused by the uplifted quantity of chemical substances that warrant being concerned.

The data-independent acquisition (DIA) and data-dependent acquisition (DDA) modes are commonly applied during NTS of complex environmental samples employing LC-HRMS, which can generate MS fragmentation spectra for compound identification. Generally speaking, DDA can obtain exclusive MS^2 spectra for a limited number of precursor ions, but many potential analytes with relatively lower intensities will be ignored. To address this concern, the DIA mode was applied to generate an unbiased MS^2 spectrum. Among several DIA modes, MS^E (i.e., MS^E function in Xevo G2-XS TOF MS, Waters) mode has shown great potential in obtaining comprehensive spectrometric information of samples, which has been applied in several studies for the identification of non-target features (Samanipour et al., 2018; Ren et al., 2022). Similar to all ions fragmentation (AIF) (Naz et al., 2017; Renaud et al., 2017), MS^E mode sends all ions in MS^1 to the collision cell for fragmentation in the next cycle of scanning, which can obtain both MS^1 and MS^2 spectra. The DDA and DIA data can be complementary to each other and their combination can provide very comprehensive information for further structural elucidation (e.g., online database matching, in-silico fragmentation).

Nonetheless, technical bottlenecks remain in NTS analysis for environmental samples. One difficulty is the heavy burden of data processing tasks because the exponentially increased data size requires sophisticated professional data processing software with an appropriate algorithm (Hohrenk et al., 2020). Many open-source and commercial software programs were developed for automated data processing, such as XCMS (Smith et al., 2006), MS-DIAL (Tsugawa et al., 2015), MZmine2 (Pluskal et al., 2010), mass profiler professional (Agilent Technologies, U.S.), Progenesis QI (Waters, U.S.) and Compound Discover (Thermo Fisher, U.S.), etc. To the best of our knowledge, none of any software can process both DDA and DIA data, and link the results together. In addition, although many studies have successfully applied these software programs to investigate the unknown contaminants in a variety of aquatic environments (such as rivers (Eysseric et al., 2021), lakes (Yao et al., 2022), marine waters (Feng et al., 2022; Lara-Martín et al., 2020), and wastewater treatment plants (Alygizakis et al., 2019; Jeong et al., 2022)), a comprehensive assessment of NTS data processing method is still lacking. This lack of assessment undermined the confidence of the NTS result (Fisher et al., 2022; Peter et al., 2021), and some researchers have remained skeptical about the reproducibility of this method (Hites and Jobst, 2018). Thus, there is an urgent need for a comprehensive assessment of the performance of the data processing method to gain confidence in this state-of-the-art approach.

To address these concerns, we have evaluated the performance of the whole NTS workflow using a self-developed program PyHRMS. This program supports users in inputting customized database files, avoiding the high cost and limited autonomy when compelled to use commercial databases. The testing compounds in this study are 38 glucocorticoids, which are one group of steroid hormones that were frequently detected in the aquatic environment and attract the attention of environmental researchers and public interest (Weizel et al., 2018; Willi et al., 2019). Peak picking efficiency, MS^2 matching performance, and quantification accuracy will be evaluated by these compounds, and a case study using this NTS workflow for a real water sample was also carried out. Identification results suggested that DIA and DDA methods are suitable for matching different kinds of compounds based on their properties, and the combination of them can maximize the screening efficiency for samples of limited prior information. The result of this study explores the boundaries of the NTS approach and will greatly improve confidence in the environmental contaminants monitoring results by NTS.

2. Materials and methods

2.1. Reagents

Glucocorticoids mix solution in acetonitrile/methanol (1:1) (Table S1), isotopically-labeled internal standards (ISTD) (Table S2), and 66 CECs for matrix spike analysis were purchased from Alta Scientific Co. (Tianjin, China). Formic acid at HPLC grade (99 %) was obtained from DiKMA Technologies (CA, U.S.). Ammonium acetate at LC/MS grade (99 %) was purchased from CNW Technologies (Shanghai, China). LC/MS-grade solvents, such as water, methanol, n-hexane, and acetone were purchased from Fisher Scientific (Pittsburgh, PA), and the ultra-pure water was produced by the Millipore Milli-Q system (U.S.).

2.2. Samples extraction and instrumental methods

Grab samples were collected in precleaned amber glass bottles and stored under 4 °C for use. All water samples were firstly filtered with glass fiber filters (0.45 μm, Jinteng, China), then spiked with a mix of ISTD standard solution, and loaded to solid phase extraction (SPE) cartridges (Oasis PRiME HLB, 500 mg, 6 cc, Waters, U.S.) for compounds extraction. All experiments in this study were conducted in triplicates. Details of sample collection and extraction method were described in Text S1. The extracts were concentrated at 1 mL and analyzed by a Waters ACQUITY ultra-performance liquid chromatography (UPLC) coupled to a quadrupole time-of-flight mass spectrometer (qTOF-MS, Xevo G2-XS) with an electrospray ionization source (ESI). Specific information about LC instrumental settings can be found in Text S2.

2.3. Data processing

For non-target analysis, the primary raw data files were converted to mzML format by MSConvert (Chambers et al., 2012) and then processed by PyHRMS (version 3.4, https://pypi.org/project/pyhrms/) for post-data processing. The primary parameters of PyHRMS for data processing are listed and interpreted in detail in our prior research (Wang et al., 2023). In conclusion, Chromatographic peaks were generated by the peak picking function in MS^1 data and were then prioritized based on the following criteria: peak area > 500, S/N > 3, fold change (vs. solvent and lab blanks) >5, and $p < 0.05$ (vs. solvent and lab blanks). The MS^2 information for these prioritized compounds was obtained by data-independent acquisition (DIA) and data-dependent acquisition (DDA) modes. A detailed description of these methods is available in Text S3. The prioritized compounds were then matched by retention time (RT) and MS^2 fragments of accurate mass information with an in-house or online database (e.g., Massbank (MassBank | MassBank of North America Mass Spectral DataBase, 2022), NORMAN database (NORMAN Suspect List Exchange (NORMAN-SLE), 2022). The confidence level of

identification based on Schymanski et al. (Schymanski et al., 2014a) was assigned to each of these compounds. Level 1 identifications were achieved by matching the RT, MS^1, and MS^2 with reference standards. Level 2 was assigned by matching major fragment ions with the online MS^2 database. The compounds that failed to match the MS^2 database can be predicted by in silico fragmentation tools (such as MetFrag (Ruttkies et al., 2016) and SIRIUS (Dührkop et al., 2019)), which were assigned to level 3. For compounds without MS^2 fragments, but their mass and isotope distribution can be matched in the merged NORMAN Suspect List Exchange Database, level 4 was assigned. The rest compounds with only the mass of interest and RT were assigned to level 5.

Identified features at level 1 confidence were then semi-quantified by a PyHRMS built-in function based on the relative response factors (RRF) method (Eq. 1–3). Each compound was assigned an internal standard (Table S1) based on their chromatographic similarities (Malm et al., 2021). The accuracy of 38 target compounds in matrix samples at 10 ng/L and 100 ng/L was evaluated based on calculated results.

$$\text{Response Factor (RF)} = \text{Peak area}/\text{Concentration} \tag{1}$$

$$\text{Relative Response Factor (RRF)} = \text{RF (target compound)} / \text{RF (reference standard)} \tag{2}$$

$$\text{Concentration A} = (\text{peak area A}/\text{peak area B}) \times (1/\text{RRF}) \times (\text{Concentration B}) \tag{3}$$

2.4. Degradation experiment

The photo-degradation experiment was conducted with flumethasone as the target compound to explore the parent-product relationships based on the diagnostic fragment ion (DFI) strategy. 500 mL 500 μg/L flumethasone water solution (with 0.1 % methanol to avoid the solvent effect) was added to a photoreactor equipped with a glass cold trap to assure light uniformity. A circulating water bath was used for reaction temperature constancy. The degradation was conducted with a xenon lamp at 1000 W/m^2 power and lasted for 30 h to simulate the transformation under natural sunlight, then the reaction solution was extracted by the method described above. Control samples were run identically in dark conditions.

2.5. Quality control and quality assurance

Before instrumental analysis, the detector performance was tuned using sodium formate, and the mass error remains <1 ppm. Leu-enkephalin (LE) (*m/z* at 556.2771 and 554.2615 for ESI^+ and ESI^-, respectively) as a calibration compound was injected into the ion source every 15 s with a real-time mass calibration performed during the whole analysis. Solvent blanks and isotope-labeled internal standard solution (Table S2) were analyzed for every 9 samples. For isotope labeled standards, if RT error > 0.05 min, mass error > 10 ppm, or relative standard deviation of peak areas >30 %, the instrument was retuned. Solvent blanks (pure methanol) and lab blanks (DI water) were analyzed alongside samples (n = 3). The peaks appeared in both samples and blanks (fold change <5 or p-values >0.05) were removed.

3. Result and discussion

3.1. Mass accuracy for glucocorticoid standards

The mass resolutions for traditional time-of-flight (TOF) mass spectrometers are usually in the range of 10,000–40,000 FWHM. There are two acquisition modes to record the raw data, i.e., profile mode and centroid mode. For profile mode, mass peaks were formed by continuous waves, while for centroid mode, the peaks were changed to bars. To obtain complete information on a mass spectrum, we selected profile mode to record the data. However, for most data processing tools, the mass value of a peak was obtained by reading the mass of its highest point, which might be inaccurate and unstable sometimes (Fig. S1). PyHRMS can optimize the mass by calculating the middle point for the full width at half the maximum of a mass peak. Fig. S2 shows the errors of both the observed and optimized mass of 38 glucocorticoids at various concentrations between 1 and 500 ng/mL. It illustrated a convincing result with mass error within ±5 ppm for optimized mass despite their various instrument responses. In contrast, the observed mass error was much bigger than the optimized mass error and can reach 10 ppm, especially for compounds of lower intensity. The mass error of repeatedly injected ISTD compounds before and after mass optimization was also summarized in Fig. S3, and most outliers were calibrated to ±5 ppm by our mass optimization function. Thus, the mass optimization by PyHRMS exhibits good performance and can provide a more precise and

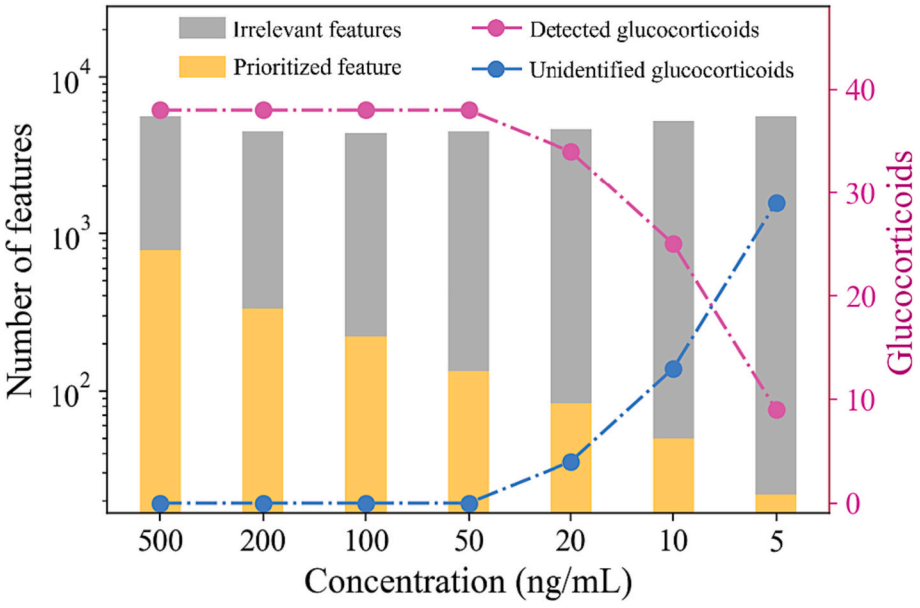


Fig. 1. Number of prioritized and irrelevant features screened out by peak picking function. Pink dots represent the number of detected glucocorticoids, and blue dots represent the number of unidentified glucocorticoids.

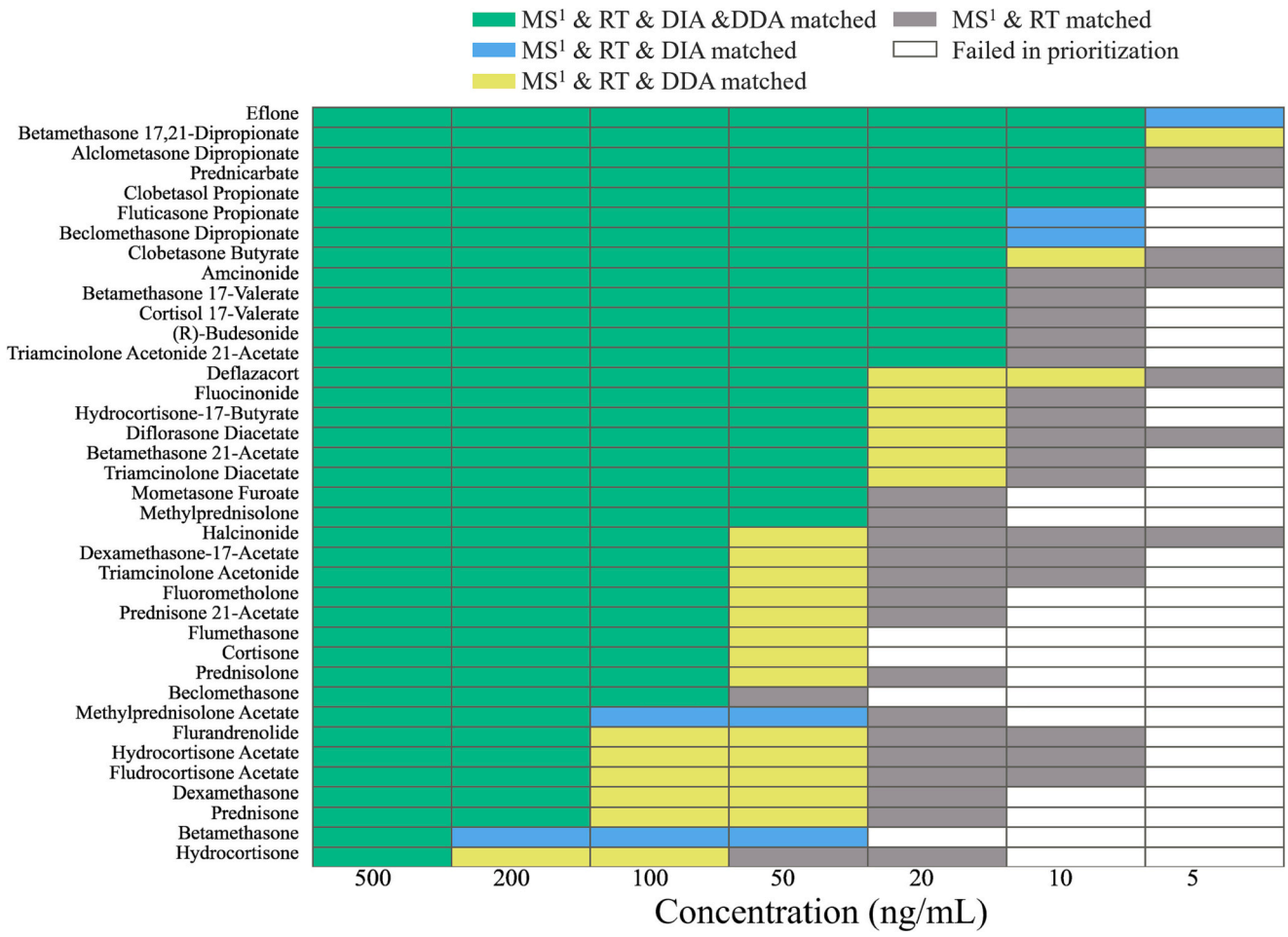


Fig. 2. In-house database matching result for glucocorticoids in dilution samples.

convincing matching result for NTS.

3.2. NTS performance for dilution curve of glucocorticoids standard methanol solution

A dilution curve of glucocorticoid standard methanol solution (5, 10, 20, 50, 100, 200, and 500 ng/mL) was used to evaluate the performance of peak picking and DDA/DIA analysis. To focus on the peak picking algorithm without interference by other factors, the recoveries for the solid extraction method and matrix effect were not considered in this case but will be investigated later.

3.2.1. Peak picking efficiency

The performance of the peak picking function for glucocorticoids in a pure solvent system was first evaluated by analyzing a standard dilution curve prepared in triplicates following the steps described in the method section. As shown in Fig. 1, the numbers of possible organic compounds in standard curve samples with S/N > 3 and intensity >500 are in the range of 4375–5646. When fold change and *p*-value filters were applied (i.e., fold change >5 and *p*-value <0.05 compared with solvent blanks), the number of prioritized features was reduced to 22–787, which accounts for 0.4 %–13.9 % of total organic compounds in each of the samples. By compounds, we mean the specific compounds composed of all possible forms of their own, including isotopes and salt adduct forms, so this number was not overestimated. As the concentration decreased, the number of prioritized compounds also decreased, while the number of detected glucocorticoids (pink dot in Fig. 1) start to decrease when the concentration was lower than 50 ng/mL. The blue dots in Fig. 1 exhibited the number of glucocorticoids that failed to be identified at different concentrations, and the reason for their misidentification was mainly the decreased peak area not meeting the criteria of over 500. In total, 220 detected peaks were grouped to true positive results by manually checking, while the missing of 46 target peaks at lower concentrations were confirmed attributing to their faint instrumental responses and inadequate chromatographic shapes.

Specific standard samples were plotted in Fig. S4, where grey dots were irrelevant compounds that were removed by fold change and *p*-value filter, while the red dots and green dots represented glucocorticoid compounds and other prioritized compounds, respectively. The other prioritized compounds might be the impurities present in the glucocorticoid stock solution, so their number also decreased as the standard concentration decreased. The decreased number of glucocorticoids in samples with a concentration lower than 50 ng/mL was mainly attributed to the unmet criteria of peak area (>500), fold change (>5), S/N (>3), etc. For example, at 10 ng/mL, betamethasone failed to be screened out by the NTS approach because its peak area was only 309 (Fig. S5). This result is reasonable and acceptable for the NTS approach, because the filter is very important to remove irrelevant or false positive features in this process, and the parameters (i.e., thresholds of *p*-value, fold change, intensity, etc.) should be adjusted according to the NTS result of a calibration curve to get the best performance and thus built a solid foundation for the following identification process.

3.2.2. DDA and DIA data analysis

The structural identification of prioritized compounds can be performed by analyzing the MS^2 information generated by collision-

induced dissociation (CID) (Horai et al., 2010; Stein, 2012). DDA and DIA are the two popular acquisition modes to acquire MS^2 information during NTS of complex samples using LC-HRMS (Guo and Huan, 2020), which can greatly reduce the labor of information input for target MS^2 acquisition. For DDA mode, we selected the first 7 highest mass peaks to obtain their MS^2 spectra for analysis. For DIA mode, we selected the all ions fragmentation (AIF) approach to make sure that enough data points were collected for chromatograms from both MS^1 and MS^2 (>22 points for each chromatographic peak) with relatively moderate scanning speed (5 Hz). The fragmentation information for AIF-based DIA was obtained by using the peak picking function for MS^2 data, and the chromatographic peaks with certain retention time and accurate mass were assigned to the prioritized compounds in MS^1 data. For example, the MS^2 spectrum of hydrocortisone acetate from DDA shows that it has fragments of *m/z* 309.1859, 327.1978, and 241.1591, etc. (Fig. S6a), while the DIA data also shows the extracted chromatographic peaks at these fragment mass (Fig. S6b). The consistency of MS^2 information from DDA and DIA indicates both methods are suitable for further in-house or online database matching.

To evaluate the performance of database matching for DDA and DIA methods, we have built an in-house database with compounds' accurate mass, retention time, and MS^2 fragments (Table S1) generated by detecting the authentic standards at 1000 ng/mL concentration with the same instrument and setting parameters described in the method section. The matching of prioritized compounds with the database must meet the following criteria: 1) retention time difference < 0.1 min; 2) precursor mass error < 10 ppm; 3) fragment ion mass error < 0.015 Da; 4) number for matched fragment ion ≥2. As shown in Fig. 2, the majority of glucocorticoid compounds in dilution curve samples can be matched well with the in-house database at relatively high concentrations (>50 ng/mL). This is the ideal situation where all RT, DDA, and DIA results can be matched with the in-house database (green rectangle in Fig. 2). However, as the concentration decreased, RT and one of the DDA/DIA data were matched with the database (yellow or blue rectangle in Fig. 2). This result is reasonable because the weak signal of glucocorticoids at low concentrations might not be enough to rank in the first 7 highest peaks in MS^1 for DDA, or cannot generate a decent chromatographic peak of fragments for DIA. As the concentration further decreased, these compounds were either matched by RT only or even not screened out by the peak picking function.

As to the MS^2 matching performance of compounds with low concentrations, both DDA and DIA have their own advantages over the others. For example, at 200 ng/mL, the betamethasone (RT: 16.4 min, *m/z*: 393.2069 ($M + H^+$)) failed to obtain its fragments by DDA (Fig. 2), because its MS^1 intensity only ranked 14th in MS^1 scan (Fig. S7a). In this case, DIA demonstrates superior performance in fragment analysis, as it can produce decent chromatographic peaks for its fragments (Fig. S7b). Conversely, in the case of dexamethasone, DDA can produce high-quality MS^2 spectrums (rank 4th highest peak in MS^1, Fig. S8a) while DIA failed in this regard because these fragments do not exhibit adequate chromatographic peaks for their faint instrument responses (Fig. S8b).

For both DDA and DIA analysis, false positive is the major concern regarding the quality of database matching results. The structural identification based on MS^2 (either DIA or DDA) is powerless for isomers. For example, dexamethasone-17-acetate (InChIKey: AKUJBENLRBOFTD-RPRRAYFGSA-N) and betamethasone 21-acetate (InChIKey: AKUJBENLRBOFTD-QZIXMDIESA-N) are two stereoisomers that have almost the same fragments (Fig. S9). In this case, if one of them was present in the samples and matched with our database, it will return two results, which need additional information for structural identification (e.g., RT or ion mobility, etc.). In addition to isomers, the background noise of mass spectra will also increase the possibility of false positives. Although DDA can generate exclusive MS^2 fragments, the background noises need to be removed by an intensity threshold filter (>300 used in this study). In contrast, the DIA data can solve this problem by peak picking in the MS^2 domain, because most noise mass cannot generate a decent chromatographic peak at the same RT with the parent prioritized compounds. In addition, DIA can obtain MS^1 and MS^2 data simultaneously with enough data points for chromatographic peaks, which exhibit a great potential to replace DDA and minimize the instrumental acquisition work in the future.

3.2.3. Online and in-house database matching

As database building and development is tedious, costly, and time-consuming work, most environmental researchers relied on the online database for the identification of unknown contaminants. The comparison between online and in-house databases is important because it can increase the reliability of spectrum information when sharing them among different labs or different LC-HRMS instruments. Compared with the spectra from GC-EI-MS, the spectra from LC-MS are less reproducible (Hollender et al., 2017b), especially when different collision energies were applied. Thus, some online databases (such as Norman-Network) only recorded the accurate fragment ion mass but omitted their relative abundance. In this scenario, we have compiled the fragment ion mass of compounds from two popular databases (i.e., Massbank of North America (MassBank | MassBank of North America Mass Spectral DataBase, 2022) and Norman network database (NORMAN Suspect List Exchange (NORMAN-SLE), 2022)) to check the MS^2 matching performance of the online database.

A total of 33 glucocorticoids were present in both online and in-house databases. As shown in Figs. S10 and S11, the performance of MS^2 matching for online and in-house are very similar in general but still different at low concentrations. This small difference is mainly attributed to the different number of fragments collected in different databases and will not affect the matching accuracy, and this result indicated that our in-house database is reliable and can be shared with other labs for MS^2 matching when an authentic standard was absent.

3.3. NTS performance for glucocorticoid compounds in matrix spike samples

3.3.1. Peak picking and MS^2 matching in matrix spike samples

Matrix spikes were prepared by spiking glucocorticoid standards in 1 L Pearl River water (10 and 100 ng/L). The samples were concentrated in 1 mL methanol solution (named matrix-10 and matrix-100, respectively) by solid phase extraction method as described in the method section. As shown in Fig. S12a, >7000 features were found in matrix-100, 2625 of them are prioritized compounds, which include all 38 glucocorticoid compounds. 30 glucocorticoid compounds were screened out in matrix-10 (Fig. S12b), which is even slightly higher than that in standard methanol solution at 10 ng/mL (i.e., 25 glucocorticoid compounds in standard-10). For fragments matching, we found that 36 glucocorticoid compounds were successfully matched by RT and fragments (either one of DDA/DIA or both) in matrix-100 (Fig. S13) versus 38 in standard-100 (Fig. 2). Similarly, 4 glucocorticoid compounds were successfully matched by RT and fragments in matrix-10 (Fig. S13) versus 8 in standard-10 (Fig. 2). This difference in NTS performance in matrix samples and pure methanol samples can be explained by the combination of solid phase extraction recoveries (40 %–150 % at 10 ng/L in ultrapure water, Fig. S14) and suppression of compounds' ionization in the ion source by the inherently existed matrix. Overall, despite the matrix effect, our NTS approach still exhibits great power for identifying the organic compounds in complex real water samples.

3.3.2. Semi-quantification for glucocorticoids by UPLC-HRMS

Quantification of prioritized contaminants identified by NTS is necessary for evaluating the site eco-risks and serving environmental management. Using HRMS data is an efficient approach but can only provide semi-quantification results for its limitations in detecting sensitivity and accuracy. Owning to limited prior information about the testing samples, incorporating isotopic-labeled internal standards for all

components in advance is unfeasible, thus, may result in inaccuracies in quantitative outcomes. Meanwhile, the inherent high background noise and narrow linear range of standard curve in TOF-MS instruments may also attribute to the uncertainty in quantification. Here, by adding selected internal standards of distinct chromatographic behaviors in real water samples, we have evaluated the semi-quantification accuracy of glucocorticoid compounds based on LC-HRMS using the relative response factor (RRF) method. The relative standard deviations of RRFs for most glucocorticoid compounds were below 35 % except for prednicarbate (RRF: 51 %) (Fig. S15), which shows an acceptable quality of internal standard curve for further semi-quantification practices. As shown in Fig. 3a, all 38 glucocorticoid compounds were screened out and the accuracies of 66 % of glucocorticoid compounds in matrix-100 are in the range of 80 %–120 %. However, in the matrix-10 sample, only 30 glucocorticoid compounds were screened out, and the highest deviation of semi-quantification result can reach to 378 % in the case of methylprednisolone (Fig. 3b). To explore the reason for this high deviation, we have obtained the extracted ion chromatogram of methylprednisolone (*m/z*: 359.1851) and sulfadimethoxine-D6 (ISTD, m/z: 317.1190). As shown in Fig. S16, the peak areas of methylprednisolone are *2577* in matrix-10 versus *917* in standard-10, and the peak areas of sulfadimethoxine-D6 in this matrix-10 and standard-10 are *34,297* and *64,345*, respectively. The enlarged peak area of methylprednisolone and the suppressed signal of ISTD led to the overestimated concentration (37.8 ng/L).

This example exposed the major flaws of the quantification by the NTS approach based on LC-HRMS data, especially for compounds at low concentrations in a complex sample (e.g., 10 ng/L methylprednisolone in real water samples). Generally speaking, the high background noise generated in LC-HRMS is an essential problem because it can interfere with the integration of peak area, which is usually overestimated. In addition to background noise, the ISTD selected for quantified compounds in NTS workflow is usually not their own isotope labeled compounds, and thus cannot accurately reflect the real recoveries and matrix effect. Despite these flaws, we are still satisfied with the result of this semi-quantification method, because even the most deviated results remained in the same order of magnitude as the true values. From a realistic perspective, this semi-quantification method exhibited great potential in future application, and the result is acceptable to provide insights into the occurrence of eco-risks of prioritized organic compounds.

3.4. Parent-product relationships identified by diagnostic fragment ion strategy

Some CECs in water can easily undergo environmental transformation processes in natural or engineering systems to generate transformation products (TPs) (Qu et al., 2013). These TPs may possess equivalent or even stronger toxicity than the parent compounds (Lao et al., 2022; Tian et al., 2022, 2021), and thus attract additional concerns. The MS^2 spectra of TPs may share the same fragments with their parent compounds because environmental transformations do not necessarily drastically alter their molecular structures. However, the structure elucidation for these TPs cannot rely on the MS2 matching because they are most likely not included in the in-house or online databases. Nonetheless, we can apply the diagnostic fragment ion (DFI) strategy to build a relationship between TPs and parent compounds based on their MS^2 spectra from DDA, and this strategy has been successfully applied in TPs identification for specific organic compounds (Wang et al., 2020; Yang et al., 2023a).

Hormones are one kind of such CECs that are easily transformed into TPs and escape from traditional monitoring (Pflug et al., 2017; Wammer et al., 2016; Yost et al., 2014). Here, we select flumethasone (InChIKey: WXURHACBFYSXBI-GQKYHHCASA-N) as a target contaminant for artificial photolytic experiment and explore its TPs by DFI strategy in NTS workflow. As shown in Fig. 4a, 505 features were prioritized by the non-target screening process for photo-degraded samples at 30 h. Those features were absent in blanks (lab blanks and solvent blanks) and control samples (same reaction system at time 0 h), indicating they are potential TPs of flumethasone. Based on a limited number of MS^2 spectra of these prioritized compounds from DDA data, a total of 13 prioritized compounds were recognized as TPs, which have two or more fragments in common with parent flumethasone. For example, the MS^2 spectrum of TP391 has *m/z* 171.0804, 235.1120, 261.1247, 275.1432, and 289.1557, etc., which are also present in the MS^2 spectrum of parent flumethasone (Fig. 4b). The mass difference between TP391 and flumethasone is 20.0084 Da, which is very close to the mass of HF (m/z 20.0062). This further verified that TP391 is the TP of flumethasone by elimination of HF from its main skeleton. For prioritized compounds that were not successfully connected with parent compounds by DFI, most of them failed to obtain their MS^2 spectra due to low intensities, and some of them cannot match with any possible formula with elements: C, H, O, N, P, Cl, F. These may be the TPs of standard impurities or preexist

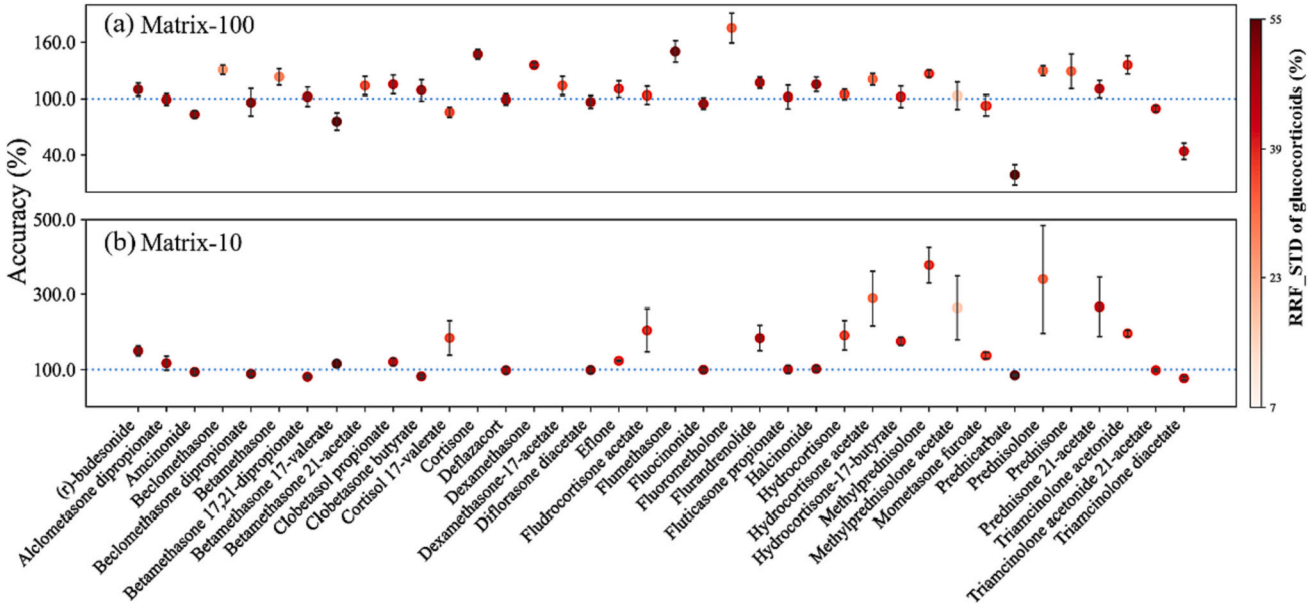


Fig. 3. HRMS quantification result of 38 glucocorticoids in matrix spiked samples at 100 ng/L (a) and 10 ng/L (b). The color of the dots represents the standard deviation of their RRF referred to as the color bar.

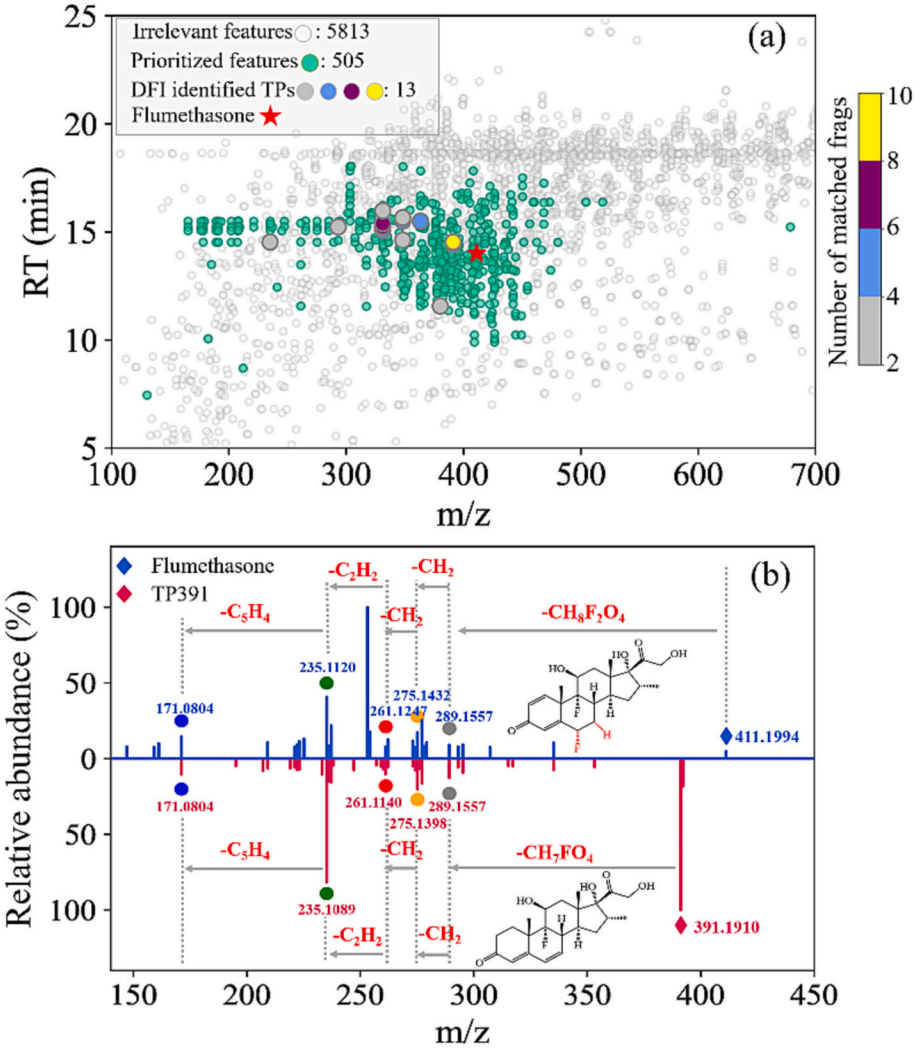


Fig. 4. (a) Mass versus RT plot of features in flumethasone photo-degraded sample (30*h*). Possible TPs identified by the DFI strategy with different numbers of matched fragments were marked in grey, blue, purple, and yellow circles, respectively; (b) comparison of MS^2 spectra (from DDA) between flumethasone and TP391. The matched fragments were marked with dots in different colors.

organic compounds in reaction systems (Yang et al., 2023a). Due to the high uncertainties for those compounds, we didn't consider these compounds as valid TPs of flumethasone. The major transformation pathways of flumethasone were summarized in Fig. S17, and we believe the DFI strategy developed in this case is very effective for identifying potential TPs of known compounds and is very helpful for building a parent-TPs relationship between unknown compounds in real water samples.

3.5. NTS case study in actual water samples

3.5.1. Matrix spike analysis in different matrices

Since glucocorticoid compounds are constrained to a narrow interval of log Kow and may exhibit similar chromatographic and mass spectrometric behaviors to some extent, we selected 66 species of commonly used chemicals in the fields of pesticides, pharmaceuticals, veterinary medicines, and industrial materials for further validation of the applicability of NTS workflow in CECs identification. These compounds exhibited a diverse instrumental retention time range of 5.3 to 21.3 min, and log Kow values spanning from −2.4 to 6.7 (Fig. S18), reflecting differentiated physical and chemical properties. Detailed information on these compounds is available in Table S3. Matrix-spiked samples were prepared using wastewater (WW), seawater (SW), and surface water (SuW) matrices that represented typical background matrices, as illustrated by their total ion chromatograms in Fig. S19. Meanwhile, all spiked CECs were confirmed to be absent in these matrices.

Fig. S20 and S21 present the identification results of spiked CECs across three dosage gradients (50, 100, and 500 ng/L), following the same HRMS data processing process as previously described. At 50 ng/L spiking dosage, >75 % of CECs were successfully prioritized in all matrix samples, demonstrating the consistent screening capability of the applied workflow. Increment in CEC dosage further aids compound prioritization to over 80 % and 95 % at 100 and 500 ng/L, respectively. Challenges in prioritization primarily arose from the inability to meet peak-picking and filtering criteria at lower dosage levels, as exemplified by dimetridazole with S/N below 3 at 50 ng/L dosage (Fig. S22). At escalated spiking dosages, CECs' prioritization efficiency improved as their instrumental response increased, with matrix-induced background noise in the spectrometer being suppressed to a negligible level. Meanwhile, approximately 50 % of CECs were identified at level 1 confidence at 50 ng/L dosage, increasing to 80 % at 500 ng/L dosage. Though the combined fragment-matching approach for both DIA and DDA data maximizes CEC identification performance, the gap between feature prioritization and CEC identification that stems from the requirement for fragment recognition remains apparent, particularly at lower dosages. Despite the screening overlooks in CECs that own low instrument

response and high matrix background, the applied NTS workflow is able to consistently deliver reliable results in feature prioritization and CEC identification across a wide concentration range under varied matrix conditions, underscoring its capability to provide convincing NTS outcomes in practical applications.

3.5.2. CECs identification in Pearl River samples by NTS methodology

To explore the application of the whole NTS workflow for real waters, one water sample was collected in the Pearl River (22°56′45.6″ N, 114°4′3.7″ E Guangdong, China) and processed by the abovementioned method. After peak picking and prioritization, 1071 and 868 features were screened out as characterized compounds in positive and negative modes, respectively. Fig. 5 shows that 18 and 1 compounds were identified by authentic standards (level 1) under positive and negative modes (Fig. S23), respectively. These compounds include pesticides, pharmaceuticals, personal care products, etc. (Sobus et al., 2018) A complete list of level 1 and level 2 compounds is shown in Table S4 (ESI^+) and S5 (ESI^-). In addition to those matched by RT, 141 compounds (including both positive and negative modes) were matched by online databases (i.e., Massbank (MassBank | MassBank of North America Mass Spectral DataBase, 2022) and NORMAN (NORMAN Suspect List Exchange (NORMAN-SLE), 2022)), which were assigned to level 2. However, there are many cases where more than one result was returned by MS^2 matching. For example, the compound with *m/z* 273.1853 at RT 18.13 min matched with two candidates from Massbank (i.e., estradiol and galaxolidone in Fig. 6a and b, respectively). Although their structures are much different, the MS^2 spectra of these two compounds shared at least 7 fragments in common, so the MS^2 data from either DDA or DIA (Fig. 6c) all matched with these two candidates. This is a good example to demonstrate that database matching may not return an exclusive result, a false positive is possible.

For compounds with MS^2 spectra from DDA but failed to match with the database, their structures can be predicted by the DFI strategy or in-silico fragmentation method. For example, we applied the DFI strategy and successfully located two compounds that are structurally related to estradiol (level 2). As shown in Fig. S24, the RT of TP273 (17.97 min) is very close to that of estradiol, and their MS^2 spectra are very similar, indicating they might be stereoisomers. Evidence in previous studies indicates that estradiol can easily transform into its isomer products in natural systems (Yang et al., 2023b). In addition, TP259 and estradiol share two fragments in common (i.e., *m/z* 129.0691 and 157.1028), and their mass difference (m/z 14.0144) is very close to the mass of CH_2 (m/z 14.0157), indicating TP259 might be a transformation product of estradiol with a CH_2 loss. The structures of these two compounds were predicted by MetFrag and are shown in Fig. S24. However, these pieces of evidence are insufficient to propose an exact structure (e.g., positional isomers), and thus the structure of these compounds remains speculative at level 3. For compounds without MS^2 spectra, level 4 was assigned if their mass was matched in Norman Suspect List (Norman Network, 2018), and the rest of the prioritized compounds were assigned to level 5.

The concentrations of compounds with level 1 are shown in Fig. S25, three compounds (i.e., 4-acetamidoantipyrine, irbesartan, and 4-Formylaminoantipyrine) are above 100 ng/L, which are all pharmaceuticals. The risk quotients (RQ) of these compounds can be calculated accordingly based on these concentrations (if their lowest Predicted No Effect Concentrations (PNEC) exist (NORMAN Ecotoxicology Database, 2022) and their environmental risks can be evaluated. Further work is also necessary to upgrade level 2 or 3 compounds to level 1 and more analysis can be done regarding the CEC characterization and occurrence at different sites with different sampling times.

This example presents a complete NTS result based on LC-HRMS, 19 compounds were confirmed with reference standards, and their concentration was obtained. However, the compounds with confidence level 1 and level 2 only account for 13.4 % of total characterized compounds, thus continuous retrospective analysis is necessary for the

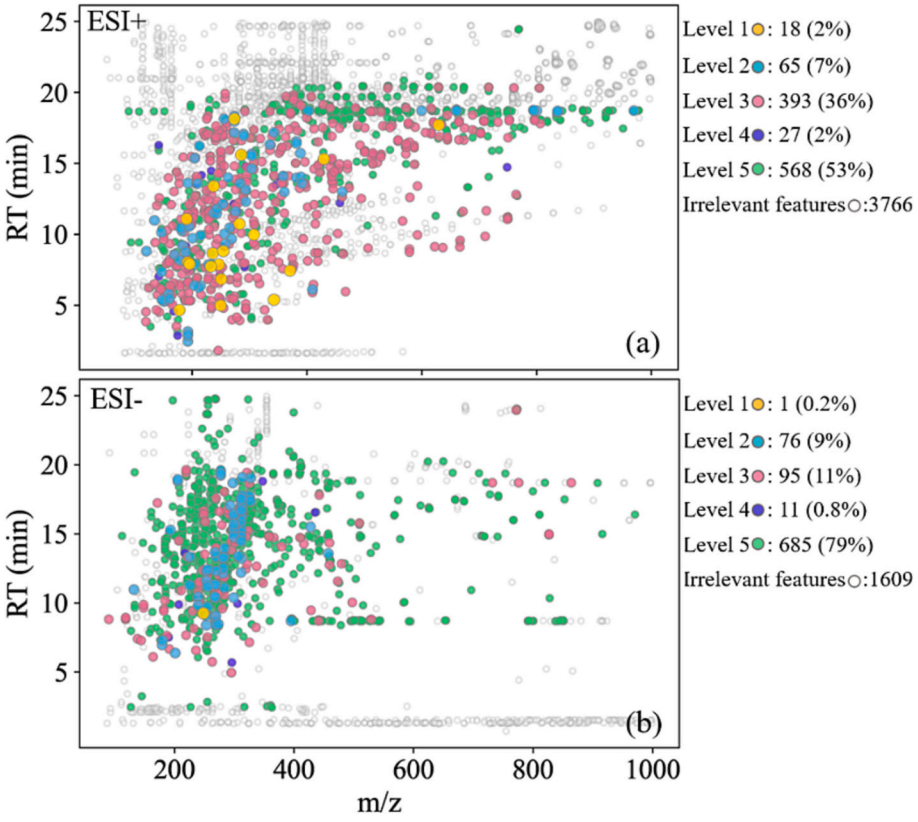


Fig. 5. Mass versus RT plots of features from the Pearl River sample under (a) positive and (b) negative modes. Dots in different colors represent the confidence level of tentatively identified features.

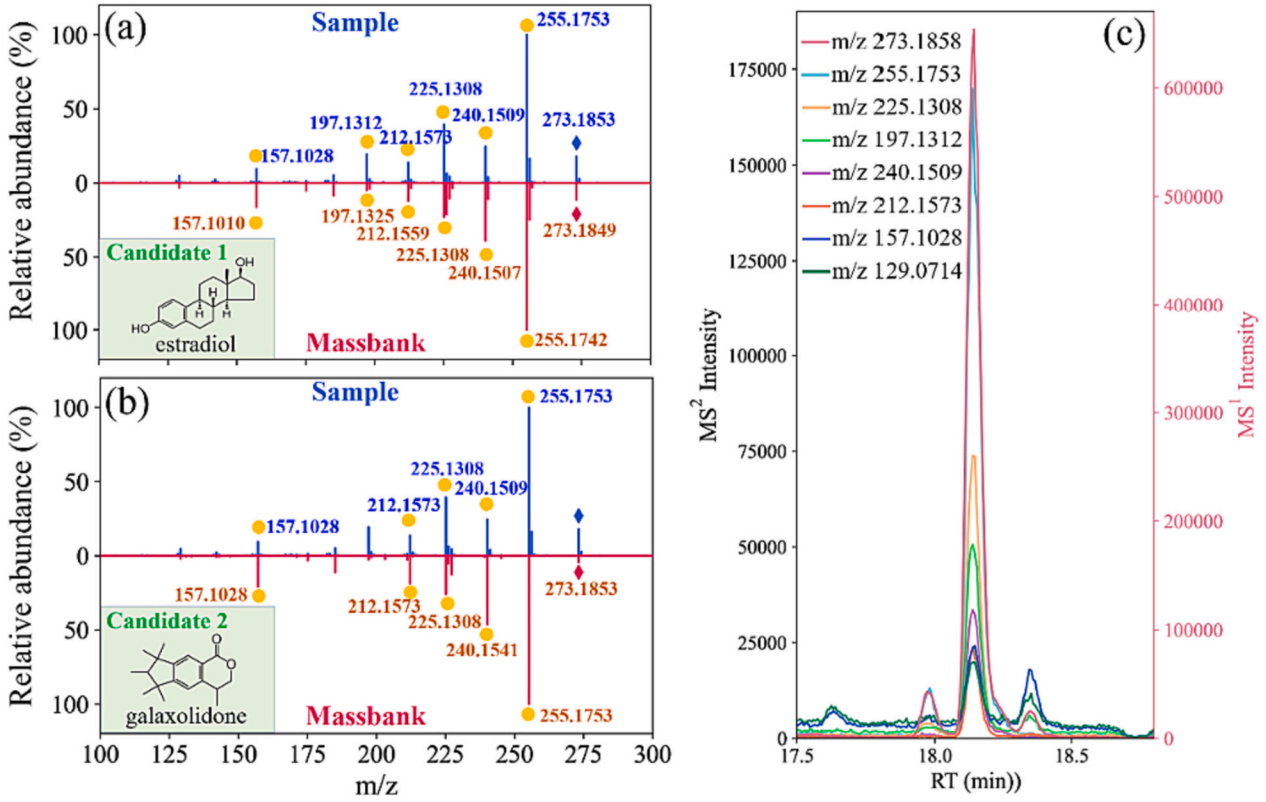


Fig. 6. MS^2 spectra comparison between DDA data (m/z 273.1853, RT 18.13) and Massbank data of two candidates (i.e., (a) estradiol and (b) galaxolidone); (c) Extracted chromatogram of matched fragments from DIA data.

future. In addition, the NTS based on LC-HRMS is not an ultimate and universal approach to characterize the whole contaminants in water, especially for compounds that are not ESI amendable (e.g., polycyclic aromatic hydrocarbon, polychlorinated biphenyls, dioxins, etc.), other extraction (e.g., liquid-liquid extraction) and analysis (e.g., GC-HRMS) methods are also needed.

4. Conclusion

This study has explored the boundaries of the NTS approach for the identification and quantification of CECs in an aquatic environment based on LC-HRMS data. Overall, the NTS workflow developed in this study was shown to be effective in the identification of unknown compounds and quantification of known compounds. This study also demonstrated the usefulness and applicability of DDA and DIA modes as well as the in-house and online database for NTS of complex environmental samples. DIA mode based on all ion fragmentation (AIF) and DDA mode exhibit comparable abilities for MS^2 matching with database, despite they have their own advantages over the other in some cases. For compound quantification, although targeted analysis by LC-MS/MS remains the most reliable approach, it is not feasible to extend this traditional quantification approach over more than a few hundred compounds. The quantification method developed in this study can be a convenient and alternative approach for the quantification of contaminants that were identified by the previous non-target screening process. In addition, the data processing algorithm in PyHRMS can process the data from the instruments of different brands, which greatly increase the cross-platform compatibility, and the evaluation of NTS performance for the different instrument will be the subject of our future studies.

CRediT authorship contribution statement

H.L.-study execution, data analysis, and manuscript writing. R.W.-study design, PyHRMS development, manuscript editing. B.Z.-study design, data analysis review and manuscript editing. D.X.- study design, data analysis review and manuscript editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This project was supported by the Fundamental Research Funds for the Central Public Welfare Research Institutes (Grant No. PM-zx703-202104-112, PM-zx703-202011-321, PM-zx703-202112-332, PM-zx703-202204-112) and Guangdong Basic and Applied Basic Research Fund (No. 2021A1515110551).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.167967.

References

Alygizakis, N.A., Samanipour, S., Hollender, J., Ibáñez, M., Kaserzon, S., Kokkali, V., van Leerdam, J.A., Mueller, J.F., Pijnappels, M., Reid, M.J., Schymanski, E.L., Slobodnik, J., Thomaidis, N.S., Thomas, K.V., 2018. Exploring the potential of a global emerging contaminant early warning Network through the use of retrospective Suspect screening with high-resolution mass spectrometry. Environ. Sci. Technol. 52, 5135–5144. https://doi.org/10.1021/acs.est.8b00365.

Alygizakis, N.A., Gago-Ferrero, P., Hollender, J., Thomaidis, N.S., 2019. Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high

fluctuation in influent wastewater. J. Hazard. Mater. 361, 19–29. https://doi.org/10.1016/j.jhazmat.2018.08.073.
Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P., 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920. https://doi.org/10.1038/nbt.2377.
Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., Böcker, S., 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat. Methods 16, 299–302. https://doi.org/10.1038/s41592-019-0344-8.
Eysseric, E., Beaudry, F., Gagnon, C., Segura, P.A., 2021. Non-targeted screening of trace organic contaminants in surface waters by a multi-tool approach based on combinatorial analysis of tandem mass spectra and open access databases. Talanta 230, 122293. https://doi.org/10.1016/j.talanta.2021.122293.
Feng, X., Sun, H., Liu, X., Zhu, B., Liang, W., Ruan, T., Jiang, G., 2022. Occurrence and ecological impact of chemical mixtures in a Semiclosed Sea by Suspect screening analysis. Environ. Sci. Technol. 56, 10681–10690. https://doi.org/10.1021/acs.est.2c00966.
Fisher, C.M., Peter, K.T., Newton, S.R., Schaub, A.J., Sobus, J.R., 2022. Approaches for assessing performance of high-resolution mass spectrometry–based non-targeted analysis methods. Anal. Bioanal. Chem. 414, 6455–6471. https://doi.org/10.1007/s00216-022-04203-3.
Günthardt, B.F., Wettstein, F.E., Hollender, J., Singer, H., Härri, J., Scheringer, M., Hungerbühler, K., Bucheli, T.D., 2021. Retrospective HRMS screening and dedicated target analysis reveal a wide exposure to pyrrolizidine alkaloids in small streams. Environ. Sci. Technol. 55, 1036–1044. https://doi.org/10.1021/acs.est.0c06411.
Guo, J., Huan, T., 2020. Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics. Anal. Chem. 92, 8072–8080. https://doi.org/10.1021/acs.analchem.9b05135.
Hites, R.A., Jobst, K.J., 2018. Is nontargeted screening reproducible? Environ. Sci. Technol. 52, 11975–11976. https://doi.org/10.1021/acs.est.8b05671.
Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of software tools for liquid chromatography–high-resolution mass spectrometry data processing in nontarget screening of environmental samples. Anal. Chem. 92, 1898–1907. https://doi.org/10.1021/acs.analchem.9b04095.
Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017a. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184.
Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017b. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ. Sci. Technol. 51, 11505–11512. https://doi.org/10.1021/acs.est.7b02184.
Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, Kenichi, Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, Ken, Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. 45, 703–714. https://doi.org/10.1002/jms.1777.
Jeong, Y., Da Silva, K.M., Iturrospe, E., Fuiji, Y., Boogaerts, T., van Nuijs, A.L.N., Koelmel, J., Covaci, A., 2022. Occurrence and contamination profile of legacy and emerging per- and polyfluoroalkyl substances (PFAS) in Belgian wastewater using target, suspect and non-target screening approaches. J. Hazard. Mater. 437, 129378 https://doi.org/10.1016/j.jhazmat.2022.129378.
Lao, J.-Y., Lin, H., Qin, X., Ruan, Y., Leung, K.M.Y., Zeng, E.Y., Lam, P.K.S., 2022. Insights into the atmospheric persistence, transformation, and health implications of organophosphate esters in urban ambient air. Environ. Sci. Technol. 56, 12003–12013. https://doi.org/10.1021/acs.est.2c01161.
Lara-Martín, P.A., Chiaia-Hernández, A.C., Biel-Maeso, M., Baena-Nogueras, R.M., Hollender, J., 2020. Tracing urban wastewater contaminants into the Atlantic Ocean by nontarget screening. Environ. Sci. Technol. 54, 3996–4005. https://doi.org/10.1021/acs.est.9b06114.
Malm, L., Palm, E., Souihi, A., Plassmann, M., Liigand, J., Kruve, A., 2021. Guide to semi-quantitative non-targeted screening using LC/ESI/HRMS. Molecules 26, 3524. https://doi.org/10.3390/molecules26123524.
MassBank | MassBank of North America Mass Spectral DataBase, 2022. MassBank. https://mona.fiehnlab.ucdavis.edu. Accessed 15-12-2022.
Naz, S., Gallart-Ayala, H., Reinke, S.N., Mathon, C., Blankley, R., Chaleckis, R., Wheelock, C.E., 2017. Development of a liquid chromatography–high resolution mass spectrometry metabolomics method with high specificity for metabolite identification using all ion fragmentation acquisition. Anal. Chem. 89, 7933–7942. https://doi.org/10.1021/acs.analchem.7b00925.
NORMAN Ecotoxicology Database, 2022. NORMAN. https://www.norman635network.com/nds/ecotox/lowestPnecIndex. Accessed 15-07-2022.
Norman Network, Aalizadeh, R., Alygizakis, N., Schymanski, E., Slobodnik, J., 2018. S0 | SUSDAT | 619 Merged NORMAN Suspect List: SusDat.
NORMAN Suspect List Exchange (NORMAN-SLE), 2022. NORMAN. https://www.norman-network.com/nds/SLE/. Accessed 15-07-2022.
Peter, K.T., Phillips, A.L., Knolhoff, A.M., Gardinali, P.R., Manzano, C.A., Miller, K.E., Pristner, M., Sabourin, L., Sumarah, M.W., Warth, B., Sobus, J.R., 2021. Nontargeted analysis study reporting tool: a framework to improve research transparency and reproducibility. Anal. Chem. 93, 13870–13879. https://doi.org/10.1021/acs.analchem.1c02621.
Petrie, B., Barden, R., Kasprzyk-Hordern, B., 2015. A review on emerging contaminants in wastewaters and the environment: current knowledge, understudied areas and recommendations for future monitoring. Water Res. 72, 3–27. https://doi.org/10.1016/j.watres.2014.08.053.
Pflug, N.C., Hankard, M.K., Berg, S.M., O'Connor, M., Gloer, J.B., Kolodziej, E.P., Cwiertny, D.M., Wammer, K.H., 2017. Environmental photochemistry of dienogest: phototransformation to estrogenic products and increased environmental persistence *via* reversible photohydration. Environ Sci Process Impacts 19, 1414–1426. https://doi.org/10.1039/C7EM00346C.
Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11, 395. https://doi.org/10.1186/1471-2105-11-395.
Qu, S., Kolodziej, E.P., Long, S.A., Gloer, J.B., Patterson, E.V., Baltrusaitis, J., Jones, G.D., Benchetler, P.V., Cole, E.A., Kimbrough, K.C., Tarnoff, M.D., Cwiertny, D.M., 2013. Product-to-parent reversion of Trenbolone: unrecognized risks for endocrine disruption. Science 342, 347–351. https://doi.org/10.1126/science.1243192.
Ren, J., Fernando, S., Hopke, P.K., Holsen, T.M., Crimmins, B.S., 2022. Suspect screening and nontargeted analysis of per- and Polyfluoroalkyl substances in a Lake Ontario food web. Environ. Sci. Technol. 56, 17626–17634. https://doi.org/10.1021/acs.est.2c04321.
Renaud, J.B., Sabourin, L., Topp, E., Sumarah, M.W., 2017. Spectral counting approach to measure selectivity of high-resolution LC–MS methods for environmental analysis. Anal. Chem. 89, 2747–2754. https://doi.org/10.1021/acs.analchem.6b03475.
Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Chemother. 8, 3. https://doi.org/10.1186/s13321-016-0115-9.
Samanipour, S., Reid, M.J., Bæk, K., Thomas, K.V., 2018. Combining a deconvolution and a universal library search algorithm for the nontarget analysis of data-independent acquisition mode liquid chromatography−high-resolution mass spectrometry results. Environ. Sci. Technol. 52, 4694–4701. https://doi.org/10.1021/acs.est.8b00259.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014a. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48, 2097–2098. https://doi.org/10.1021/es5002105.
Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014b. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environ. Sci. Technol. 48, 1811–1818. https://doi.org/10.1021/es4044374.
Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem. 78, 779–787. https://doi.org/10.1021/ac051437y.
Sobus, J.R., Wambaugh, J.F., Isaacs, K.K., Williams, A.J., McEachran, A.D., Richard, A. M., Grulke, C.M., Ulrich, E.M., Rager, J.E., Strynar, M.J., Newton, S.R., 2018. Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. J. Expo. Sci. Environ. Epidemiol. 28, 411–426. https://doi.org/10.1038/s41370-017-0012-y.
Stein, S., 2012. Mass spectral reference libraries: an ever-expanding resource for chemical identification. Anal. Chem. 84, 7274–7282. https://doi.org/10.1021/ac301205z.
Tian, Z., Peter, K.T., Gipe, A.D., Zhao, H., Hou, F., Wark, D.A., Khangaonkar, T., Kolodziej, E.P., James, C.A., 2020. Suspect and nontarget screening for contaminants of emerging concern in an urban estuary. Environ. Sci. Technol. 54, 889–901. https://doi.org/10.1021/acs.est.9b06126.
Tian, Z., Zhao, H., Peter, K.T., Gonzalez, M., Wetzel, J., Wu, C., Hu, X., Prat, J., Mudrock, E., Hettinger, R., Cortina, A.E., Biswas, R.G., Kock, F.V.C., Soong, R., Jenne, A., Du, B., Hou, F., He, H., Lundeen, R., Gilbreath, A., Sutton, R., Scholz, N.L., Davis, J.W., Dodd, M.C., Simpson, A., McIntyre, J.K., Kolodziej, E.P., 2021. A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. Science 371, 185–189. https://doi.org/10.1126/science.abd6951.
Tian, Z., Gonzalez, M., Rideout, C.A., Zhao, H.N., Hu, X., Wetzel, J., Mudrock, E., James, C.A., McIntyre, J.K., Kolodziej, E.P., 2022. 6PPD-Quinone: revised toxicity assessment and quantification with a commercial standard. Environ. Sci. Technol. Lett. 9, 140–146. https://doi.org/10.1021/acs.estlett.1c00910.
Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., Arita, M., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat. Methods 12, 523–526. https://doi.org/10.1038/nmeth.3393.
Wammer, K.H., Anderson, K.C., Erickson, P.R., Kliegman, S., Moffatt, M.E., Berg, S.M., Heitzman, J.A., Pflug, N.C., McNeill, K., Martinovic-Weigelt, D., Abagyan, R., Cwiertny, D.M., Kolodziej, E.P., 2016. Environmental photochemistry of Altrenogest: Photoisomerization to a bioactive product with increased environmental persistence via reversible Photohydration. Environ. Sci. Technol. 50, 7480–7488. https://doi.org/10.1021/acs.est.6b02608.
Wang, R., Yan, Y., Liu, H., Li, Yanxi, Jin, M., Li, Yuqing, Tao, R., Chen, Q., Wang, X., Zhao, B., Xie, D., 2023. Integrating data dependent and data independent non-target screening methods for monitoring emerging contaminants in the Pearl River of Guangdong Province, China. Sci. Total Environ. 164445 https://doi.org/10.1016/j.scitotenv.2023.164445.
Wang, X., Yu, N., Yang, J., Jin, L., Guo, H., Shi, W., Zhang, X., Yang, L., Yu, H., Wei, S., 2020. Suspect and non-target screening of pesticides and pharmaceuticals

transformation products in wastewater using QTOF-MS. Environ. Int. 137, 105599 https://doi.org/10.1016/j.envint.2020.105599.
Weizel, A., Schlüsener, M.P., Dierkes, G., Ternes, T.A., 2018. Occurrence of glucocorticoids, mineralocorticoids, and progestogens in various treated wastewater, Rivers, and streams. Environ. Sci. Technol. 52, 5296–5307. https://doi.org/10.1021/acs.est.7b06147.
Willi, R.A., Salgueiro-González, N., Carcaiso, G., Fent, K., 2019. Glucocorticoid mixtures of fluticasone propionate, triamcinolone acetonide and clobetasol propionate induce additive effects in zebrafish embryos. J. Hazard. Mater. 374, 101–109. https://doi.org/10.1016/j.jhazmat.2019.04.023.
Yang, X., Wang, R., He, Z., Dai, X., Jiang, X., Liu, H., Li, Y., 2023a. Abiotic transformation of synthetic progestins in representative soil mineral suspensions. J. Environ. Sci. 127, 375–388. https://doi.org/10.1016/j.jes.2022.06.007.
Yang, X., Wang, R., He, Z., Dai, X., Jiang, X., Liu, H., Li, Y., 2023b. Abiotic transformation of synthetic progestins in representative soil mineral suspensions. J. Environ. Sci. 127, 375–388. https://doi.org/10.1016/j.jes.2022.06.007.
Yao, J., Sheng, N., Guo, Y., Yeung, L.W.Y., Dai, J., Pan, Y., 2022. Nontargeted identification and temporal trends of per- and Polyfluoroalkyl substances in a Fluorochemical industrial zone and adjacent Taihu Lake. Environ. Sci. Technol. 56, 7986–7996. https://doi.org/10.1021/acs.est.2c00891.
Yost, E.E., Meyer, M.T., Dietze, J.E., Williams, C.M., Worley-Davis, L., Lee, B., Kullman, S.W., 2014. Transport of steroid hormones, phytoestrogens, and estrogenic activity across a swine lagoon/Sprayfield system. Environ. Sci. Technol. 48, 11600–11609. https://doi.org/10.1021/es5025806.