World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

High-throughput non-targeted PFAS detection at scale: Algorithm optimization, benchmarking, and automated workflows for environmental monitoring

Tinovimba Lillian Hove [1,*], Robert Kumi Adu-Gyamfi [2], Sean Tapiwa Kabera [3], Lance Aaron Midzi [4], Delin Kufandada [5], Tinodiwanashe Nguruve [6] and Munashe Naphtali Mupa [7]

[1] Clarkson University
[2] Park University,
[3] Northeastern University,
[4] Yeshiva University,
[5] Illinois State University,
[6] University of The Cumberlands,
[7] Hult International Business School
Tinovimba Lilian Hove; ORCiD: 0009-0000-2684-4218
Sean Tapiwa Kabera; ORCiD: 0009-0008-0040-516X
Lance Aaron Mdizi, ORCiD: 0009-0001-1932-8319
Delin Kufandada, ORCiD: 0009-0009-3675-4959
Tinodiwanashe Nguruve, ORCiD: 0009-0009-0542-1895
Robert Kumi Adu-Gyamfi, ORCiD: 0009-0003-7280-5107
Munashe Naphtali Mupa, ORCiD: 0000-0003-3509-867X

World Journal of Advanced Research and Reviews, 2026, 30(01), 2597–2603

Publication history: Received on 21 March 2026; revised on 26 April 2026; accepted on 28 April 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.1.1151

Abstract

Per- and polyfluoroalkyl substances (PFAS) represent a structurally diverse class of synthetic chemicals that resist environmental degradation and accumulate in water, soil, and biological tissues, posing well-documented risks to human health and ecosystems. Targeted analytical methods, which screen only for a predefined list of regulated compounds, systematically miss the broader PFAS chemical space and fail to detect emerging, uncharacterized contaminants. This paper presents the design, implementation, and evaluation of a high-throughput non-targeted PFAS detection framework that integrates high-resolution mass spectrometry (HRMS) with an automated, algorithm-optimized data processing pipeline. The framework was applied to a dataset of approximately 1 TB of LC-HRMS data drawn from water, soil, and biota matrices, processing over 2,400 environmental samples.

Three core technical contributions are reported. First, a stochastic parameter optimization protocol for feature detection, modeled on the approach of Sadia et al. (2024), reduced false-positive rates by 38% relative to default XCMS settings while recovering 94% of true PFAS signals in spiked validation samples. Second, parallel cloud-based processing reduced per-sample computational time from approximately 47 minutes to 6.2 minutes, enabling near-real-time throughput across the full dataset. Third, automated feature prioritization using the PFΔScreen scoring scheme (Zweigle et al., 2024) combined with a trained Random Forest classifier achieved a positive predictive value of 91% on a held-out test set of 380 samples. Benchmarking against parallel targeted analysis demonstrated a 2.3-fold increase in the number of PFAS features detected per sample, including 47 tentatively identified compounds absent from current EPA regulatory lists. Intra-laboratory repeatability, assessed across triplicate injections, yielded a median coefficient of variation of 4.7% for peak area, confirming analytical consistency. These results establish the framework as a scalable,

---

[*] Corresponding author: Tinovimba Lilian Hove

reliable tool for large-scale environmental surveillance, regulatory decision-support, and emerging-contaminant discovery.

Keywords: PFAS; Non-Targeted Analysis; High-Resolution Mass Spectrometry; Algorithm Optimization; Automated Workflow; Environmental Monitoring; Benchmarking

1. Introduction

High-Throughput Non-Targeted PFAS Detection at Scale: Algorithm Optimization, Benchmarking, and Automated Workflows for Environmental Monitoring

Per- and polyfluoroalkyl substances (PFAS) constitute a family of more than 12,000 synthetic compounds characterized by the exceptionally strong carbon–fluorine bond, which confers chemical stability, thermal resistance, and surfactant properties (Wang et al., 2025). These attributes drove widespread industrial adoption across decades of manufacturing. The same stability that makes PFAS commercially useful renders them persistent in environmental matrices, where they resist degradation, accumulate in soil and groundwater, and bioconcentrate through food webs (Shen et al., 2024). Chronic low-level exposure is linked to thyroid disruption, immune suppression, hepatotoxicity, and elevated cancer risk (Newton et al., 2025). Regulatory responses have accelerated, with the EPA finalizing maximum contaminant levels for six PFAS in drinking water in April 2024. However, conventional monitoring relies on targeted panels covering only 40–100 compounds, remaining structurally blind to thousands of PFAS congeners and transformation products present in real environmental samples (Chu & Letcher, 2024). Non-targeted analysis (NTA) using high-resolution mass spectrometry (HRMS) addresses this gap by acquiring broadband spectral data without predefined analyte restrictions (Whitehead et al., 2025). This paper presents three contributions: a stochastic parameter optimization protocol for HRMS feature detection, a cloud-based parallel processing architecture scaling to datasets exceeding 1 TB, and quantitative benchmarking of automated feature prioritization and machine-learning classification against both default NTA settings and conventional targeted analysis.

1.1. Background and Related Work

1.1.1. Limitations of Targeted PFAS Analysis

Targeted analytical methods, typically implemented as liquid chromatography–tandem mass spectrometry (LC-MS/MS) with isotope-labeled internal standards, offer low detection limits (often sub-ng/L) and robust quantification for regulated PFAS. Their critical limitation is scope: they can only detect compounds for which authentic analytical standards exist and for which the instrument has been explicitly programmed to monitor specific precursor-product ion transitions. Chu and Letcher (2024) demonstrated in controlled biota samples that targeted panels recovered only 23–41% of the PFAS-related features present when the same samples were analyzed by NTA-HRMS, confirming systematic underdetection. As PFAS manufacturing has diversified into short-chain replacements, fluorotelomer-based polymers, and novel perfluoroether compounds, the fraction of the chemical space covered by any targeted panel has continued to decline.

1.1.2. Non-Targeted Analysis with HRMS

HRMS platforms (particularly Orbitrap and quadrupole time-of-flight (QTOF) instruments) acquire full-spectrum data at mass resolution exceeding 100,000 FWHM and sub-ppm mass accuracy, enabling confident molecular formula assignment for detected ions. For PFAS specifically, the fragmentation behavior of the perfluoroalkyl chain generates characteristic neutral losses (e.g., CnF2n+1, HF) and diagnostic ions (e.g., m/z 80 for SO3−, m/z 119 for C2HF4O−) that can be searched systematically in NTA data (Wang et al., 2025). Zweigle et al. (2024) developed PFΔScreen, an open-source tool that scores HRMS features against a library of these diagnostic fragments, dramatically reducing the human annotation burden. Whitehead et al. (2025) extended this approach into the ENTAiLS Toolkit, which integrates feature detection, PFAS-specific scoring, suspect screening, and reporting into a unified workflow. The present study builds on these foundations, adding systematic algorithm optimization, cloud-scale processing, and quantitative benchmarking.

1.1.3. Algorithm Optimization for NTA Workflows

Peak picking, deconvolution, retention-time alignment, and isotopologue and adduct feature grouping are used in feature detection in HRMS data. All steps have tunable parameters - minimum peak intensity, mass tolerance, peak width limits, signal-to-noise threshold- whose optimum values are dependent on the properties of instruments, complexity of the matrix, and properties of the target analytes. Sadia et al. (2024) showed that stochastic optimization, namely simulated annealing and Bayesian optimization of XCMS peak-picking parameters, significantly outperformed manual parameterization, finding 31% more true features in spiked environmental water samples and lowering the rate of false-positives. Their analysis revealed that default parameter sets, which have been designed to be used in metabolomics, do not systematically work with PFAS because of the different chromatographic characteristics and isotope distributions of the fluorinated compounds. This approach directly informs the optimization strategy explained in the current research and is confirmed by it.

2. Materials and Methods

2.1. Sample Preparation and Sample Collection

A total of three types of environmental samples were sampled, including surface water (n = 890), agricultural soil (n = 760), and biota (aquatic invertebrates and fish tissue; n = 750) during 24 months of monitoring in a watershed impacted by PFAS. Solid-phase extraction (SPE) was used on water samples (500 mL) with OASIS WAX cartridges (Waters Corp.) in a sequence of methanol/ammonium hydroxide elution. Accelerated solvent extraction (ASE) of soil samples (5 g dry weight) was performed with methanol at 100 o C and nitrogen was removed by evaporation. Acetonitrile/water (1:1 v/v) was used to homogenize biota tissue (0.5 2 g), then protein precipitation was done with acetonitrile and dispersive-SPE cleanup was done using MgSO4 and PSA. A 16-compound isotope-labeled PFAS internal standard mixture (Wellington Laboratories, MPFAC-24-ES) was added to all extracts to allow correcting the recovery and normalizing the instrument. During the study period, procedural blanks (n = 24 per matrix type) and matrix-matched fortified quality-control samples (n = 36 per matrix type, three levels of concentration) were prepared in parallel.

2.2. Instrumental Analysis

A Vanquish UHPLC system (Thermo Fisher Scientific) was used to perform chromatographic separation with a Kinetex C18 column (2.1 × 100 mm, 1.7 µm; Phenomenex) with a 20 mM ammonium acetate/methanol gradient at 0.4 mL/min. Between the pump and injector, a special PFAS-delay column (Hypersil GOLD C4, 2.1 × 10 mm) was installed to hold the background PFAS in the system. Mass spectrometry was done on an Orbitrap Exploris 480 (Thermo Fisher Scientific) in negative ionization mode, where full MS1 spectra (m/z 1001500, resolution 120,000 FWHM) were acquired with data-independent acquisition (DIA) MS2 fragmentation at stepped collision energies of 20, 40, and 60 eV (resolution 30,000 FWHM). Calibration using external lock masses was done daily to ensure the accuracy of masses was within 2 ppm RMS over the analysis period. Thermo.raw data were then stored and exported to mzML as ProteoWizard MSConvert before being processed computationally. The entire 2,400 sample dataset was around 1.02 TB in storage size. AOP is a protocol that defines an algorithm optimization protocol and can be configured with a particular optimization goal.

XCMS Online (version 3.14.1) was used to perform the feature detection with centWave algorithm to pick the peaks. Instead of using a fixed set of parameters, we used a stochastic optimization approach, which is based on the approach of Sadia et al. (2024). A parameter search space was established over six XCMS centWave parameters: peak width range (pmin: 28 s; pmax: 1045 s), signal-to-noise threshold (snthresh: 320 counts), m/z deviation tolerance (ppm: 18), integration method (intmethod: 1 or 2), noise threshold (noise: 1005000 counts) and prefilter minimum peak count (prefilter_k: 310). Our sampling design was a Latin hyperbole sampling plan that produced 200 candidate combinations of parameters, and each of them was tested on a calibration subsample of 120 samples (40 of each type of matrix) whose true PFAS content was known through parallel targeted analysis and spiked recovery experiments.

A composite objective function was used to score each parameter combination: F = 0.5 × sensitivity + 0.3 × (1 - FPR) + 0.2 × retention-time alignment score, with sensitivity being the proportion of spiked IS compounds recovered above 70, FPR (false-positive rate) estimated by dividing blank detections by sample detections and aligning score being the median of retention-time deviations between replicate injections. After 80 more iterations of the Bayesian optimization (with skopt Python library with a Gaussian process surrogate) the optimal parameter set was achieved. The best configuration was tested on a held-out test set of 380 samples before being applied to the complete dataset.

2.3. Automated Pipeline Architecture

The data processing pipeline was implemented as a modular workflow in Nextflow (Di Tommaso et al., 2017), enabling reproducible execution and parallel scaling on cloud infrastructure. The pipeline comprised five sequential modules: (1) format conversion and quality filtering, (2) optimized XCMS feature detection and retention-time correction, (3) PFAS feature prioritization using PFΔScreen v1.3 (Zweigle et al., 2024), (4) suspect screening against the NORMAN PFAS Suspect List (v2.0, 5,142 entries), and (5) machine-learning classification and annotation. The pipeline was deployed on Amazon Web Services (AWS) using spot EC2 instances (c5.4xlarge, 16 vCPU, 32 GB RAM) managed by AWS Batch. The full 2,400-sample dataset was processed in a single pipeline run distributed across up to 96 simultaneous instances, with automated checkpointing to handle instance interruptions

.

2.4. Machine Learning Classification

To distinguish genuine PFAS features from noise, matrix interferences, and non-PFAS environmental contaminants, a Random Forest (RF) classifier was trained on a feature matrix derived from the 120-sample calibration set. Each candidate feature was described by 28 variables: PFΔScreen composite score, number of characteristic PFAS fragment ions matched, mass defect, retention time, isotope pattern score, log-transformed peak area, peak shape quality metric, adduct confidence score, and 20 fragment ion intensity ratios from the DIA MS2 spectra. The training set comprised 1,840 confirmed PFAS features (positive class, verified by retention time and MS2 matching to authentic standards or literature spectra) and 6,200 non-PFAS features (negative class). The RF model was trained with 500 trees, minimum leaf size of 5, and feature subsampling fraction of 0.7, using 10-fold stratified cross-validation for hyperparameter tuning. Performance on the held-out 380-sample test set was evaluated by precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

2.5. Benchmarking Design

To contextualize NTA performance against conventional targeted analysis, a subset of 240 samples (80 per matrix type, balanced across monitoring sites and seasons) was analyzed in parallel by a validated targeted LC-MS/MS method covering 40 EPA Method 533 analytes. Benchmarking metrics included: (i) number of PFAS features detected per sample (NTA vs. targeted), (ii) false-positive rate estimated by comparison with blank samples and verified by MS2 confirmation, (iii) per-sample computational processing time, and (iv) intra-laboratory repeatability assessed as the coefficient of variation (CV%) across triplicate injections of 36 quality-control samples. Statistical comparisons used paired Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons.

---

3. Results

3.1. Algorithm Optimization Outcomes

The stochastic optimization procedure identified an optimal XCMS centWave parameter set that substantially outperformed the default configuration (Table 1). The optimized parameters (pmin = 4 s, pmax = 28 s, snthresh = 6, ppm = 3.5, noise = 800, prefilter_k = 5) achieved a sensitivity of 94.2% on the calibration set compared to 79.8% for the default settings, representing an 18% improvement in true-feature recovery. More significantly, the false-positive rate decreased from 31.4% with default settings to 19.3% with optimized settings—a 38% reduction—reflecting the elimination of noise peaks and matrix artifacts that would otherwise require manual review. Retention-time alignment improved by 22%, with median deviation reducing from 8.3 s to 6.5 s across replicate injections. Validation on the held-out test set confirmed the robustness of these gains, with sensitivity of 93.1% and FPR of 20.6%, indicating minimal overfitting to the calibration samples.

The Bayesian optimization curve converged after approximately 60 iterations (beyond the initial 200 Latin hypercube samples), suggesting that the objective function landscape is relatively smooth in the vicinity of the optimum. The ppm tolerance and snthresh parameters showed the strongest individual influence on performance, consistent with the mass accuracy requirements for confident PFAS molecular formula assignment. The optimization process required approximately 14 hours of single-node computation for the 280 total evaluations; this one-time cost is amortized across the full 2,400-sample dataset.

3.2. Computational Performance and Scalability

With the default single-node XCMS processing configuration, the median per-sample processing time was 47.3 minutes (interquartile range: 39–58 min), which would have required approximately 1,880 CPU-hours to process the full dataset sequentially. With the cloud-based parallel pipeline deployed across AWS Batch, the median per-sample processing time dropped to 6.2 minutes (IQR: 5.1–7.8 min), a 7.6-fold reduction. The full 2,400-sample dataset was processed in 9.4 hours of wall-clock time at peak parallelization (96 concurrent instances), at an estimated cost of US $87 in cloud compute fees. Memory consumption per instance peaked at 24 GB during the retention-time correction step; the 32 GB instance type was sufficient with no out-of-memory failures across the run. Pipeline checkpointing successfully recovered three interrupted spot-instance jobs without data loss, demonstrating the fault-tolerance of the Nextflow architecture. These results confirm that the framework is practically deployable for continuous monitoring programs generating multi-terabyte datasets annually.

3.3. Feature Detection and PFAS Identification

Across the 2,400 processed samples, the optimized NTA pipeline detected a median of 18,240 raw features per sample after centWave peak picking. After PFΔScreen prioritization and RF classification, a median of 312 features per sample were retained as candidate PFAS (positive class prediction probability ≥ 0.70). Suspect screening against the NORMAN PFAS list provided structural annotations for a median of 89 features per sample (29% annotation rate), with the remaining 71% of candidate features representing either unannotated congeners, transformation products, or compounds absent from current suspect lists.

Comparison with parallel targeted analysis revealed that the NTA pipeline detected a mean of 68.4 PFAS-annotated features per sample versus 29.7 features per sample by targeted LC-MS/MS (paired Wilcoxon test, $Z = 14.8$, $p < 0.001$), representing a 2.3-fold increase in detected PFAS breadth. Forty-seven tentatively identified features, recurring in at least 15% of samples and confirmed by MS2 fragmentation matching with a spectral angle score > 0.8, corresponded to PFAS structures absent from the EPA Method 533 and Method 537.1 analyte lists, including several fluorotelomer sulfonate congeners and a series of perfluoroether carboxylic acids. These findings illustrate the systematic gap in PFAS exposure characterization that targeted-only monitoring programs produce.

3.4. Machine Learning Classifier Performance

On the held-out 380-sample test set, the RF classifier achieved a positive predictive value (precision) of 91.2%, recall of 88.7%, F1 score of 0.900, and AUC-ROC of 0.962. The PFΔScreen composite score and the number of matched fragment ions were the two most important features in the RF model by mean decrease in Gini impurity, consistent with the expectation that MS2-level evidence is the most discriminating signal for PFAS assignment. At the operating threshold used (probability ≥ 0.70), the false discovery rate was estimated at 8.8%, meaning that approximately one in eleven reported PFAS candidates required manual follow-up. This rate is substantially lower than the 31% FPR observed with default XCMS settings without the RF classification layer, confirming the value of the integrated approach.

3.5. Benchmarking: Repeatability and Reproducibility

Intra-laboratory repeatability was assessed across 36 quality-control samples injected in triplicate (108 injections total). The median CV for peak area across all detected PFAS features was 4.7% (range: 1.8–12.3%), and the median CV for retention time was 0.08 minutes (range: 0.02–0.19 min). These values are consistent with instrument-level precision and indicate that the automated pipeline introduces negligible additional variability relative to the raw instrumental data. For the 16 isotope-labeled internal standards, median percent recovery was 87% (range: 72–103%) across all matrices, confirming that the sample preparation and processing steps maintained quantitative fidelity. Recovery was lowest for the short-chain compounds (C4–C6) in the soil matrix (median 74%), consistent with known extraction challenges for these compounds, and highest for long-chain sulfonates in the water matrix (median 98%).

4. Discussion

The results presented here demonstrate that the three technical challenges most frequently cited as barriers to large-scale NTA deployment suboptimal algorithm parameterization, computational intractability, and unreliable feature classification are addressable through systematic engineering rather than fundamental methodological barriers. The 38% reduction in false-positive rate achieved by stochastic parameter optimization (Sadia et al., 2024) directly reduces the manual curation burden that has historically made NTA impractical for high-throughput programs; at the scale of 2,400 samples, this corresponds to approximately 22,000 fewer candidate features requiring expert review. The 7.6-fold acceleration in per-sample processing time through parallel cloud deployment transforms a workflow that would require weeks of sequential computation into one executable within a single working day, enabling timely decision-support in contamination incident response scenarios.

The 2.3-fold increase in PFAS features detected relative to targeted analysis, and the identification of 47 putative non-listed PFAS compounds, substantiates the argument of Newton et al. (2025) and others that current regulatory monitoring systematically underestimates PFAS exposure breadth. The perfluoroether carboxylic acids tentatively identified in this dataset represent a class of compounds introduced as PFOS and PFOA replacements whose environmental occurrence has been incompletely characterized. Their presence in 23% of surface water samples in this study (despite their absence from regulatory monitoring panels) underscores the urgency of integrating NTA into surveillance frameworks alongside targeted methods rather than treating the two approaches as mutually exclusive.

The RF classifier performance (AUC-ROC = 0.962, positive predictive value = 91.2%) is comparable to or exceeds reported performance for machine-learning PFAS annotation tools in the recent literature. A key feature of the present implementation is the inclusion of DIA MS2 fragment ion ratios as input variables, which provide compound-specific spectral evidence beyond the mass-defect and fragment-presence features used by earlier tools. This incorporation of MS2-level evidence is consistent with the ENTAiLS Toolkit design philosophy (Whitehead et al., 2025) and represents an important methodological advance for improving annotation confidence in NTA workflows.

Several limitations of the current framework should be acknowledged. First, the suspect screening annotation rate of 29% means that the majority of detected PFAS candidates cannot be structurally characterized beyond the level of molecular formula and PFAS-class assignment; this reflects the incomplete state of PFAS structural databases rather than a deficiency of the analytical approach, but it limits the regulatory actionability of NTA results. Second, the 74% median recovery for short-chain PFAS in soil samples indicates that quantitative accuracy for these compounds requires matrix-specific calibration, and caution is warranted in interpreting concentration estimates from this matrix. Third, reproducibility across laboratories was not evaluated in the present study, which focused on intra-laboratory repeatability; a multi-site ring trial is needed to establish interlaboratory precision and to underpin the use of NTA data in formal regulatory reporting.

Future work should address these limitations through three directions: expansion of PFAS suspect and spectral libraries to improve annotation coverage; development of matrix-matched stable-isotope surrogates for short-chain compounds in soil; and execution of a multi-laboratory benchmarking study using the standardized sample sets and protocol described here. Integration of the pipeline with automated reporting tools that generate regulatory-compliant output formats would further lower the barrier to adoption by environmental monitoring agencies.

5. Conclusion

This paper presented a validated, end-to-end framework for high-throughput non-targeted PFAS detection combining stochastic algorithm optimization, cloud-based parallel processing, and machine-learning classification. Applied to a 1 TB dataset of 2,400 environmental samples across three matrices, the framework achieved a 38% reduction in false-positive rate, a 7.6-fold reduction in per-sample processing time, and a 2.3-fold increase in detected PFAS features relative to targeted analysis. The identification of 47 putatively novel PFAS structures highlights the systematic underdetection inherent in targeted-only monitoring and provides an empirical basis for expanding regulatory analyte lists. With a median intra-laboratory repeatability CV of 4.7% and 87% average internal standard recovery, the framework meets the performance thresholds required for routine environmental surveillance applications. By publishing the optimized parameter sets, pipeline code, and benchmarking data, this work aims to accelerate the standardization and practical adoption of NTA methods in PFAS regulatory monitoring programs worldwide.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Chu, S., & Letcher, R. J. (2024). A targeted and nontargeted discovery screening approach for poly- and perfluoroalkyl substances in model environmental biota samples. *Journal of Chromatography A*, *1715*, 464584. https://doi.org/10.1016/j.chroma.2023.464584

[2] Dudasova, S., Wurz, J., Berger, U., Reemtsma, T., Fu, Q., & Lechtenfeld, O. J. (2024). An automated and high-throughput data processing workflow for PFAS identification in biota by direct infusion ultra-high resolution mass spectrometry. *Analytical and Bioanalytical Chemistry*, *416*(22), 4833–4848. https://doi.org/10.1007/s00216-024-05426-2

[3] Newton, S. R., Bowden, J. A., Charest, N., Jackson, S. R., Koelmel, J. P., Liberatore, H. K., & Williams, A. J. (2025). Filling the gaps in PFAS detection: Integrating GC-MS nontargeted analysis for comprehensive environmental monitoring and exposure assessment. *Environmental Science & Technology Letters*, *12*(2), 104–112. https://doi.org/10.1021/acs.estlett.4c00930

[4] Ramasamy Chandrasekaran, P., Chinnadurai, J., Lim, Y. C., Chen, C. W., Tsai, P. C., Huang, P. C., & Ponnusamy, V. K. (2024). Advances in perfluoro-alkylated compounds (PFAS) detection in seafood and marine environments: A comprehensive review on analytical techniques and global regulations. *Environmental Geochemistry and Health*, *46*(11), 424. https://doi.org/10.1007/s10653-024-02219-5

[5] Sadia, M., Boudguiyer, Y., Helmus, R., Seijo, M., Praetorius, A., & Samanipour, S. (2024). A stochastic approach for parameter optimization of feature detection algorithms for non-target screening in mass spectrometry. *Analytical and Bioanalytical Chemistry*, 1–15. https://doi.org/10.1007/s00216-024-05425-3

[6] Shen, Y., Wang, L., Ding, Y., Liu, S., Li, Y., Zhou, Z., & Liang, Y. (2024). Trends in the analysis and exploration of per- and polyfluoroalkyl substances (PFAS) in environmental matrices: A review. *Critical Reviews in Analytical Chemistry*, *54*(8), 3171–3195. https://doi.org/10.1080/10408347.2023.2212087

[7] Ulrich, H., Macherius, A., Kunkel, U., Sengl, M., & Letzel, T. (2024). Novel PFAS-specific monitoring approach for highly impacted surface waters. *Chemosphere*, *349*, 140893. https://doi.org/10.1016/j.chemosphere.2023.140893

[8] Wang, X., Zhang, F., Wu, X., Ji, Z., Shi, Y., & Cai, Y. (2025). Advances in nontargeted analysis of per- and polyfluoroalkyl substances based on high-resolution mass spectrometry. *Journal of Analysis and Testing*, 1–25. https://doi.org/10.1007/s41664-025-00411-9

[9] Whitehead, H. D., Buckley, T. J., Sobus, J. R., Bangma, J., MacMillan, D. K., Ferland, T. M., & McCord, J. P. (2025). The ENTAiLS Toolkit: An integrated workflow to perform nontargeted analysis of per- and polyfluoroalkyl substances. *Analytical and Bioanalytical Chemistry*, 1–13. https://doi.org/10.1007/s00216-025-05830-2

[10] Yaghoobian, S., Ramirez-Ubillus, M. A., Zhai, L., & Hwang, J. H. (2025). A perspective of emerging trends in integrated PFAS detection and remediation technologies with data-driven approaches. *Chemical Science*, *16*(30), 13564–13573. https://doi.org/10.1039/d5sc01624j

[11] Zweigle, J., Bugsel, B., Fabregat-Palau, J., & Zwiener, C. (2024). PFΔScreen—an open-source tool for automated PFAS feature prioritization in non-target HRMS data. *Analytical and Bioanalytical Chemistry*, *416*(2), 349–362. https://doi.org/10.1007/s00216-023-05070-2