Ministry of Environment of Denmark
Environmental Protection Agency

HITLIST2

Validation of a Non-Targeted Screening methodology for use in monitoring of xenobiotics in the aquatic environment

Environmental Project no. 2176

August 2021

Publisher: The Danish Environmental Protection Agency

Editors:
Emil Egede Frøkjær, Aarhus University, Department of Environmental Science
Martin Hansen, Aarhus University, Department of Environmental Science

This scientific report has been reviewed by eleven Steering Group Members from the Danish Environmental Protection Agency: Helle Rüsz Hansen, Lisa Vogel, Maj-Britt Bjergager, Janni Magelund Degn Larsen, Lisbeth Sejerøe, Gudrun Frandsen Krog, Cecilie Kjer Elkjær, Anna Gade Holm, Torben Wandall, Mia Roest Christensen, and Anna Marie Sørensen.

Graphics: Department of Environmental Science, Aarhus University

ISBN: 978-87-7038-334-9

Contents

Acronyms

(H)ESI	(Heated) electrospray ionisation
(N)CE	(Normalised) collision energy
AEC	Anion-exchange chromatography
AGC	Automatic gain control (ion population)
CD	Compound Discoverer
CEC	Cation-exchange chromatography
CFC	Chlorofluorocarbon
CI	Chemical ionisation
CLAM	Continuous low-level aquatic monitoring
$ddMS^2$	Data-dependent fragmentation
EI	Electron impact ionisation
FA	Formic acid
FWHM	Full width at half maximum
GC	Gas chromatography
HDPE	High-density polyethylene
HILIC	Hydrophilic interaction chromatography
HLB	Hydrophilic-lipophilic balanced polymer
HRMS	High-resolution mass spectrometry
HAA	Haloacetic acids
IC	Ion exchange chromatography
IT	Injection time
LC	Liquid chromatography
LLE	Liquid-liquid extraction
LOD	Limit of detection
MAX	Mixed-anion exchange
MCX	Mixed-cation exchange
MeCN	Acetonitrile
MeOH	Methanol
MFS	Environmental xenobiotics, “miljøfarlige stoffer”
MSA	Methanesulfonic acid
nLC	nano-liquid chromatography
NTS	Non-targeted screening analysis
PAH	Polycyclic aromatic hydrocarbon
PCB	Polychlorinated biphenyl
PFAS	Per- and polyfluoroalkyl substance
ppb	Parts-per-billion (ng/mL, µg/L)
ppm	Parts-per-million (µg/mL, mg/L)
QC	Quality control
RF	Radio frequency
SPE	Solid phase extraction
TF	TraceFinder
TFA	Trifluoroacetic acid
VOC	Volatile organic compound

1. Introduction

The national environmental water quality monitoring program, NOVANA, determines, amongst others, pesticide residues concentrations in the aquatic environment on a regular basis. These measures are based on highly specialised sensitive and accurate targeted analytical methods. However, a momentous drawback of targeted analytical approaches is the exclusive focus on a predefined compound list for detection. Hence, data on other chemical entities or transformation products and metabolites, potentially also present in the given sample, are not recorded, and this information is lost.

Non-targeted screening analysis (NTS) is a novel holistic approach, based on high-resolution mass spectrometry (HRMS), that rapidly profile thousands of (unknown) substances in complex environmental samples [1]–[3]. The NTS strategy is used when former unknown compounds are detected in a sample and data is investigated without any presumptions or knowledge of the sample [4]. Suspect screening is another strategy used for searching HRMS data for known chemicals, i.e. by using a reference list of pesticides and biocides which are expected to be present in the sample [4]. As such, NTS is used to describe this entire field of research.

The NTS concept was recently developed and applied in a research project under the Danish Environmental Protection Agency's Pesticide Research Program [1]. From this work it was concluded that a broader investigation of the chemical space that can be captured by NTS and validation work was needed prior to a potential implementation in national monitoring programs.

The aim of the present research project was to validate an NTS methodology, so it can be used as reliable monitoring method for e.g. the groundwater monitoring program GRUMO. By targeting 967 xenobiotic compounds (labelled as MFS, '*miljøfarlige stoffer*' or environmental pollutants of special concern), this project investigated if a wide chemical space was detected on five different state-of-the-art NTS-platforms (e.g. liquid, gas, and ion chromatography-based HRMS methods). Furthermore, sample preparation and detection limits were evaluated for every MFS.

2. The chemical space

The first aim of this project was to generate a list containing nearly 1000 MFS of interest for water quality monitoring that covers substances spanning a broad range of chemical properties. A suspect list of 967 xenobiotics was constructed to contain environmental pollutants of special concern (MFS). All chemicals are listed in Appendix 1 and their selection, based on input from the steering group and environmental chemistry and fate knowledge, is described more in detail in Appendix 2.

2.1 Chemical classes

The 967 chemicals were manually categorised and sub-divided into nine chemical classes or groups (Figure 1).

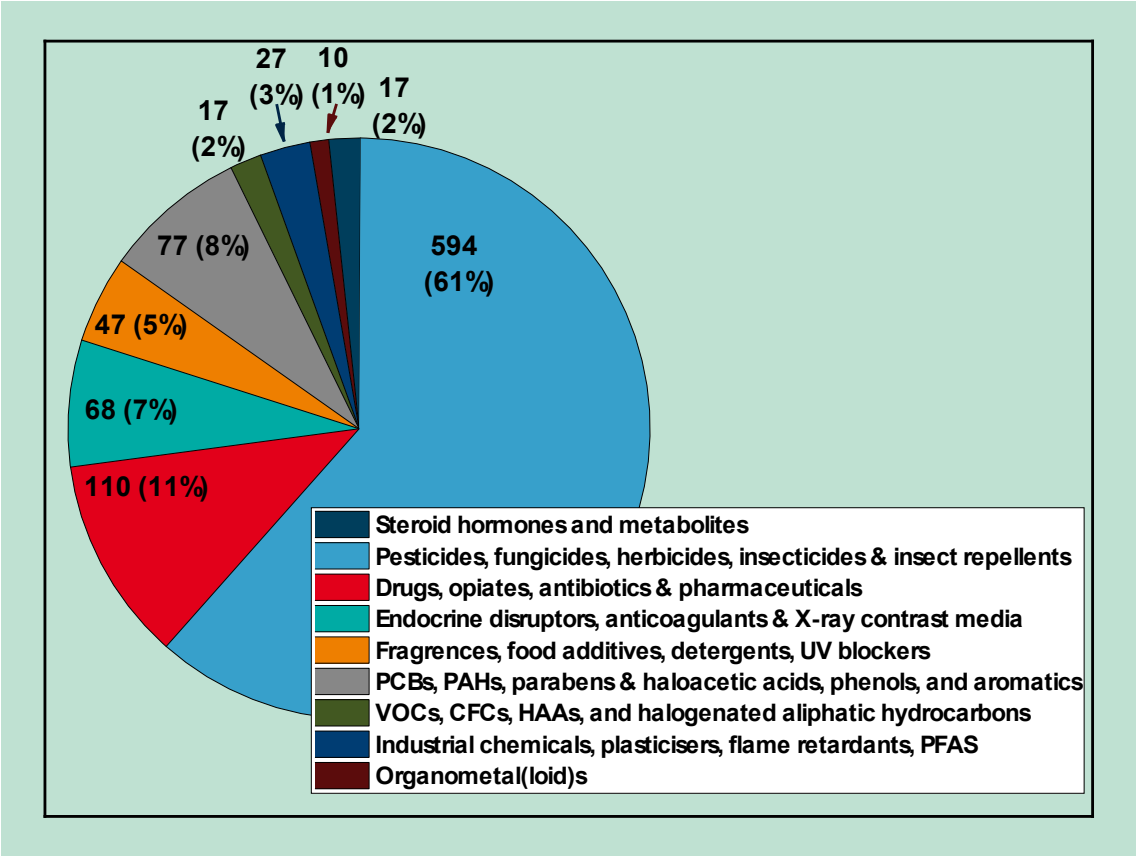

FIGURE 1: Overview of the investigated chemical classes. The number of substances is listed for each class. n = 967.

2.2 The chemical space

To investigate how well these chemical groups were resolved in the NTS platforms, structural and physiochemical information was obtained from e.g. ChemSpider, PubChem and PAN Pesticide Database using the webchem package for R (version 4.0.2) as described elsewhere [5] and curated by a workflow similar to what is described by Gadaleta et al. [6]. Despite missing data entries, general trends can still be seen throughout the dataset. The 967 compounds were selected without prior knowledge of their physiochemical properties. The resulting properties are shown in Figure 2, where the distribution of mass, logP (octanol-water partitioning coefficient),

pKa (acid strength), and volatility is shown for all compounds, in good agreement with what is previously reported in literature [7].

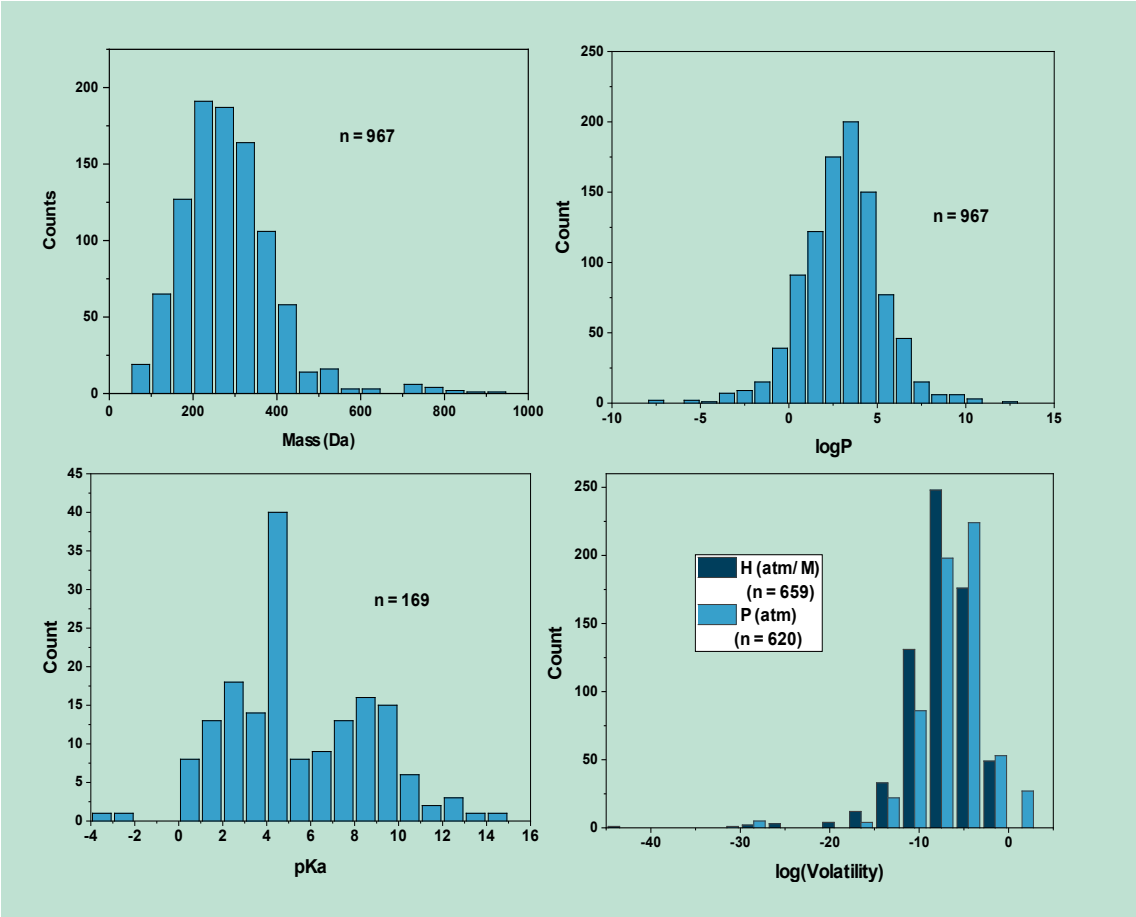

FIGURE 2: Overview of the predicted physiochemical properties of the 967 xenobiotic compounds. Top left) Molecular weight distribution (n = 967). Top right) logP distribution (n = 967). Bottom left) pKa distribution (n = 169). Bottom right) Distribution of volatility parameters H, Henry's constant (atm/M) (n = 659) and P, vapour pressure (atm) (n = 620). Both values have been adjusted by the logarithm.

2.3 Analytical platforms

To investigate the NTS-coverage of the wide chemical space, three major analytical platforms were employed: Liquid chromatography (LC) to cover mid-range polar compounds, gas chromatography (GC) for volatile non-polar compounds, and ion-exchange chromatography (IEC) for charged and ionic compounds [7]. In this project we will refer to five platforms based on different ionisation modes:

1) nano-Liquid chromatography electrospray ionisation high-resolution tandem mass spectrometry:
   I. Positive ionisation: nLC-ESI(+)-HRMS
   II. Negative Ionisation: nLC-ESI(-)-HRMS

2) Ion-exchange high-performance chromatography electrospray ionisation high-resolution tandem mass spectrometry:
   III. Anion-exchange, negative ionisation: AEC-ESI(-)-HRMS
   IV. Cation-exchange, positive ionisation: CEC-ESI(+)-HRMS

3) Gas chromatography electron impact ionisation high-resolution mass spectrometry
   V. Positive ionisation: GC-EI(+)-HRMS

3. Results and discussion

3.1 Identification of chemical standards on five NTS platforms

The 967 chemically pure analytical standards were purchased and analysed on the five analytical platforms. The number of compounds that were resolved in each respective platform is shown in Table 1. Of the 967 compounds analysed, 679 (70 %) were identified on one or more platforms. Figure 3 shows how each platform overlaps and are complementary to each other in regards to detection.

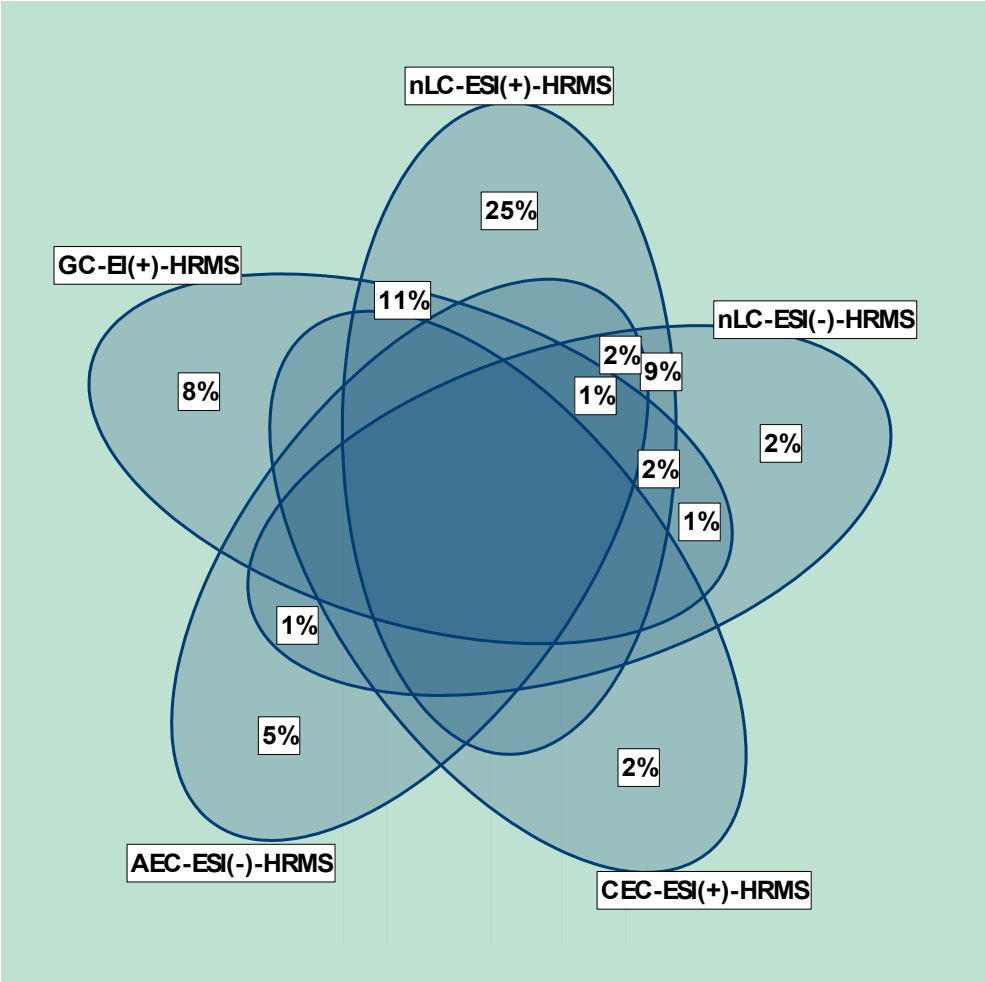

FIGURE 3: Venn-diagram of platform coverage. Values are given in percentage of identified compounds out of 967. Fields representing 0 % (< 5 compounds) are not labelled.

TABLE 1: Resolved standards on five analytical platforms.

Platform	Resolved standards	Unique for platform
nLC-ESI(+)-HRMS	492 (51%)	243 (25%)
nLC-ESI(-)-HRMS	178 (18%)	20 (2%)
CEC-ESI(+)-HRMS	19 (2%)	16 (2%)
AEC-ESI(-)-HRMS	100 (10%)	50 (5%)
GC-EI(+)-HRMS	229 (24%)	76 (8%)
Total	**679** out of **967** (70%)	

nLC-ESI(+)-HRMS is seen to resolve the highest amount of compounds (51 %) with 25 % of the compounds being entirely unique (only detected) to this platform. Combined with both AEC-ESI(-)-HRMS and GC-EI(+)-HRMS, these platforms can complement each other well, covering up to 68 % of the studied chemical space. While cation-exchange chromatography, having only a few overlapping compounds, showed the most unique annotation capability, it was also the least resolving platform.

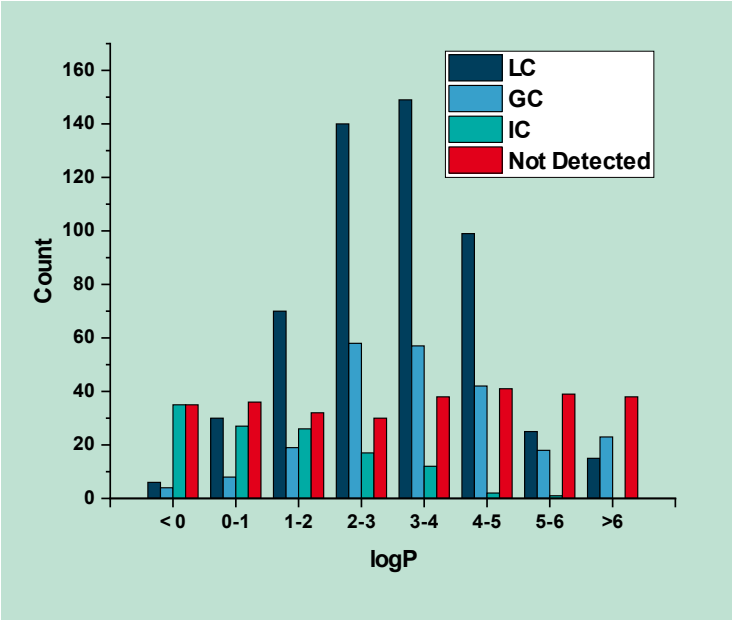

FIGURE 4: Overview of covered polarity ranges in terms of logP on each platform.

Figure 4 and Figure 5 display the distribution of logP-values of compounds identified on each platform. The measured polarity ranges for each platform shows that ion-exchange chromatography covers a larger polarity range and better resolves polar and ionic compounds than both liquid and gas chromatography. Liquid and gas chromatography generally seem to cover a similar polarity range. Anion-exchange chromatography therefore stands out as complementary technique to nLC-ESI(+)-HRMS compared to GC-EI(+)-HRMS, as this would allow for coverage of a larger chemical space in terms of polarity.

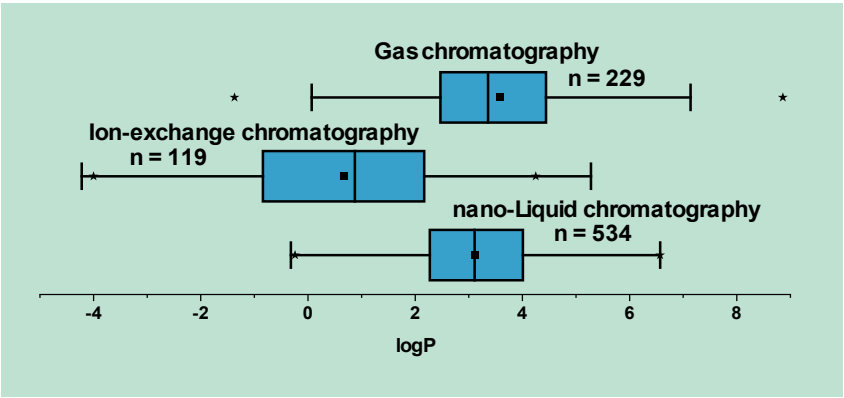

FIGURE 5: Box-plots of covered polarity ranges in terms of logP on each platform. Range given as the 5th to 95th percentile. Squares denote the mean. Markers for 1 and 99 % percentiles are given by the stars. Outliers are excluded. Overlapping compounds detected between the two ionisation modes in nano-liquid chromatography were excluded, leading to a final number of n = 534.

3.1.1.1 LC-HRMS

A relation between retention times and logP for compounds resolved in LC-HRMS can be seen in Figure 6. Such relationship can assist in the identification of unknown chemicals.

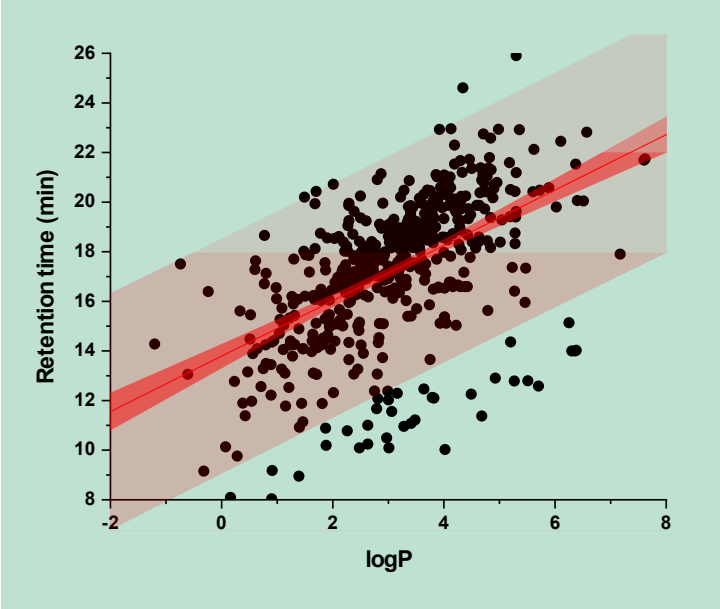

FIGURE 6: Relationship between measured retention times and estimated logP values of compounds detected on the LC-HRMS platform.

3.1.1.2 IEC-HRMS

The ion-exchange platform was able to resolve 119 of the xenobiotic compounds (cations and anions). This included the ability to resolve haloacetic acids like trifluoroacetic acid (TFA), which was detected with an estimated instrumental detection capability of 1-10 µg/mL as seen in Figure 7.

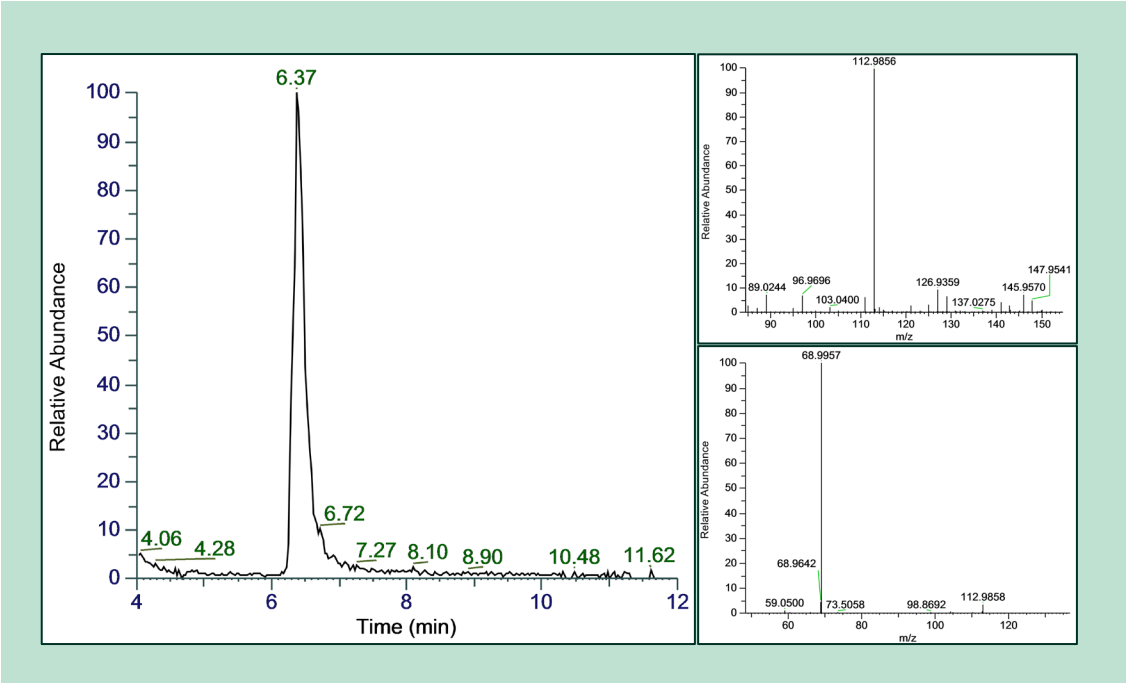

FIGURE 7: Resolution of TFA standard resolved on the AEC-ESI(-)-HRMS platform. Left) Extracted-ion chromatogram. Top right) MS1-spectrum of TFA. Bottom right) MS2-spectrum of TFA.

3.1.1.3 GC-HRMS

No clear difference in covered polarity range is seen in GC-HRMS over LC-HRMS. However highly lipophilic substances such as polychlorinated biphenyls (PCBs) and polyaromatic hydrocarbons (PAHs) were detected only with this platform.

3.2 Method validation

Sample preparation and acquisition techniques were validated by compound specific spike-recovery experiments and calibration curves. The used sample preparation method was based on enriching a 2-litre water sample via tandem solid-phase extraction and analysing the combined extract across all NTS-platforms (see section 5.5.2 for details). Estimated recoveries and detection limits are presented in Table 2. The exact recoveries for every compound can be found in Appendix 1.

TABLE 2: Overview of method validation results. Recoveries at 100 ng/L spike level in groundwater are given for both glass (top value) and plastic (bottom value) sampling containers. A confidence level of 95% were used to determine confidence intervals. Samples stored in plastic containers were not analysed by GC-EI(+)-HRMS.

Platform	Recovered in matrix	Recovery (%)	LOD (µg/L)*
nLC-ESI(+)-HRMS	427 (87%)	120 ± 7 119 ± 7	0.04 ± 0.01 (n = 412)
nLC-ESI(-)-HRMS	150 (84%)	119 ± 21 105 ± 13	0.02 ± 0.00 (n = 140)
CEC-ESI(+)-HRMS	15 (79%)	10 ± 10 10 ± 10	0.02 ± 0.01 (n = 15)
AEC-ESI(-)-HRMS	84 (84%)	32 ± 11 35 ± 12	0.03 ± 0.02 (n = 76)
GC-EI(+)-HRMS	180 (79%)	88 ± 11	0.03 ± 0.01 (n = 158)

* To avoid strong outliers, only results belonging to spiking concentrations of 1 µg/L and below have been included.

Acceptable recoveries within the range of 70 - 130 % are seen for both LC platforms and the GC platform, whereas the two IEC platforms have somewhat lower recoveries. Estimated values for matrix-matched detection limits all fall between 0.01 µg/L - 0.1 µg/L. These limits could likely be lowered by further optimization of the sample preparation or a repetition of standard addition experiment using lower reference standard concentrations. Figure 8 gives an overview of the recovered analytes on each respective platform as well as showcasing the overlaps. While nLC-ESI(+)-HRMS seems to capture the most compounds in the aquatic environment, certain compound groups are better resolved on other platforms. Despite its wide coverage of both pharmaceuticals, pesticides, organotins, and other xenobiotics, nLC-ESI(+)-HRMS fails to resolve most nitrophenols, PFAS, and certain acidic compounds, that are all better resolved by nLC-ESI(-)-HRMS. Likewise, AEC-ESI(-)-HRMS are better at resolving smaller polar compound groups such as haloacetic acids, contrast media, and polar and ionic acids. GC-EI(+)-HRMS covers a similar range to that of nLC-ESI(+)-HRMS with the exception of better selectivity towards fragrances and (semi-)volatiles, PCBs, and PAHs. For use on aqueous matrices however, it is debatable how relevant GC-EI(+)-HRMS is compared to nLC-ESI(+)-HRMS due to the hydrophobic properties and low water solubilities of the GC favourable compounds.

The largest coverage of the chemical space of xenobiotics relevant to the aquatic environment is achieved with the combination nLC-ESI(+)-HRMS, nLC-ESI(-)-HRMS, and AEC-ESI(-)-HRMS. The neglection of the GC-EI(+)-HRMS platform is based on the assumption that only small quantities of hydrophobic compounds like PCBs and PAHs are expected to be found in aquatic matrices, and would if included, only marginally increase the identifiable chemical space.

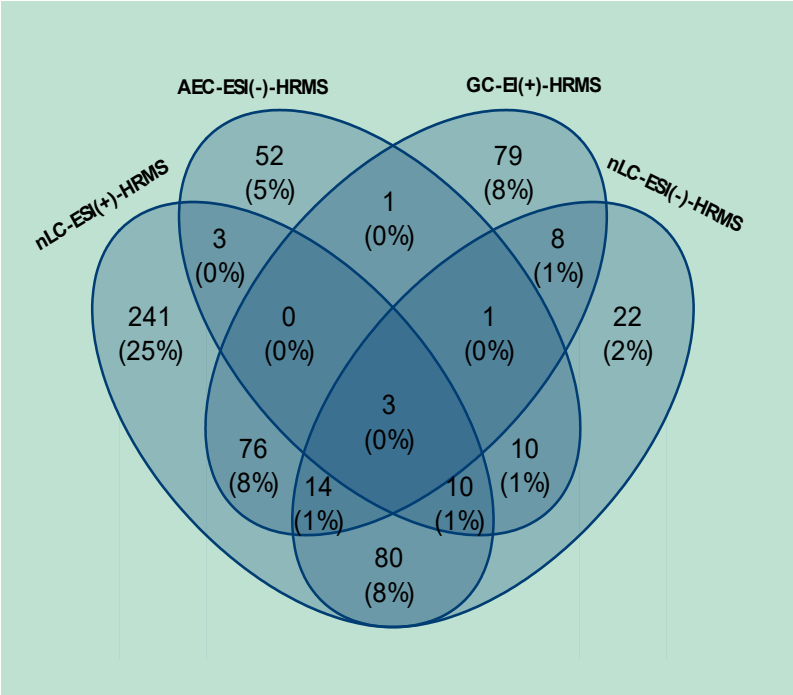

FIGURE 8: Venn diagram displaying the number of analytes recovered (out of 967) in 100 ng/L spiked groundwater samples on each respective platform. For simplicity, CEC-ESI(+)-HRMS is excluded due to the low number of recovered compounds when using this platform.

3.2.1 Recoveries

3.2.1.1 LC-HRMS

From Figure 8 we see that the combined LC-HRMS platform resolve a total of 468 unique compounds in the 0.1 µg/L post-spiked groundwater samples - 41 on nLC-ESI(-)-HRMS, 320 on nLC-ESI(+)-HRMS, and 107 overlapping compounds. Of the 178 compounds resolved on the nLC-ESI(-)-HRMS platform (Table 1), 150 were detected in the post-spiked groundwater samples. Similarly, of the 492 compounds resolved on the nLC-ESI(+)-HRMS platform, 427 were detected in the post-spiked groundwater samples. An overview of the recoveries for the detected compounds on the two LC-HRMS platforms sampled in both plastic and glass containers are shown in Figure 9.

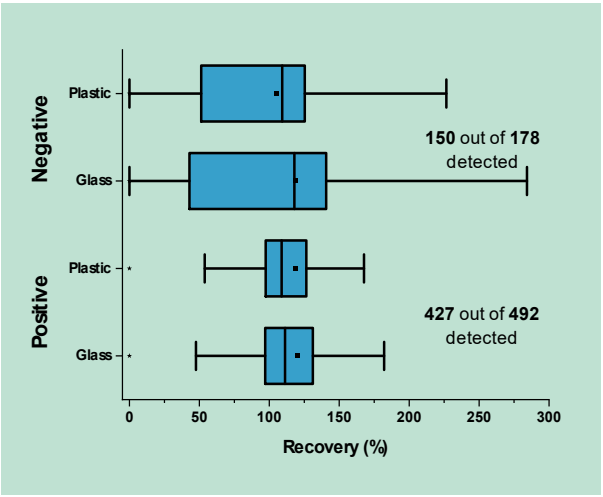

FIGURE 9: Box-plots of the recoveries of 150 (both in plastic and glass) compounds on nLC-ESI(-)-HRMS (top) and 427 compounds on nLC-ESI(+)-HRMS (bottom) in glass and plastic containers respectively.

3.2.1.2 IEC-HRMS

A combined total of 119 compounds were resolved on the cationic and anionic IEC-HRMS platforms. Of these, 99 were successfully identified in the 0.1 µg/mL post-spiked ground water samples. The recoveries for the detected compounds on the two IEC-HRMS platforms sampled in both plastic and glass containers are shown in Figure 10.

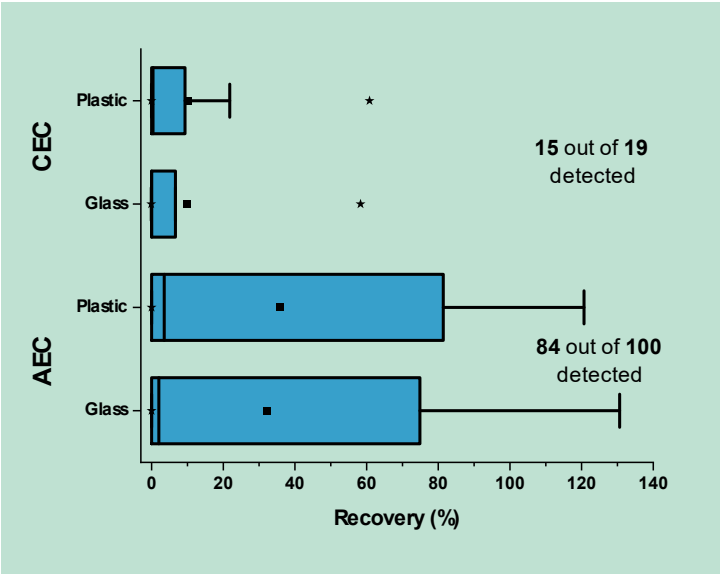

FIGURE 10: Box-plots of the recoveries of 15 compounds on CEC-ESI(+)-HRMS (top) and 84 compounds on AEC-ESI(-)-HRMS (bottom) in glass and plastic containers respectively.

3.2.1.3 GC-HRMS

A combined total of 229 compounds were resolved on the GC-ESI(+)-HRMS platform. Of these, 180 were successfully identified in the 0.1 µg/mL post-spiked ground water samples. The recoveries for the detected compounds on the GC platform sampled in glass containers are shown in Figure 11.

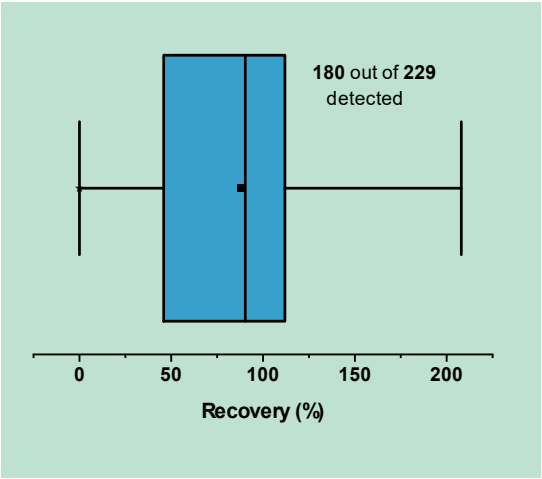

FIGURE 11: Box-plots of the recoveries of 180 compounds on GC-EI(+)-HRMS in measured in glass containers

3.2.2 Estimated detection limits

Matrix-matched post-spiked extracts of 0.1 µg/mL were used - corresponding to a sample concentration of 0.1 µg/L, when considering an enrichment factor of 1000 for the extraction of 2 L sample into a 2 mL extract.

3.2.2.1 LC-HRMS

Due to issues with the linear dynamic range on the LC-HRMS platform (see section 5.6.6), estimated detection limits were calculated using both five- and four-point calibration. The lowest of the two values were assigned as the final LOD. Estimated LOD's for the LC-HRMS platforms can be seen in Figure 12.

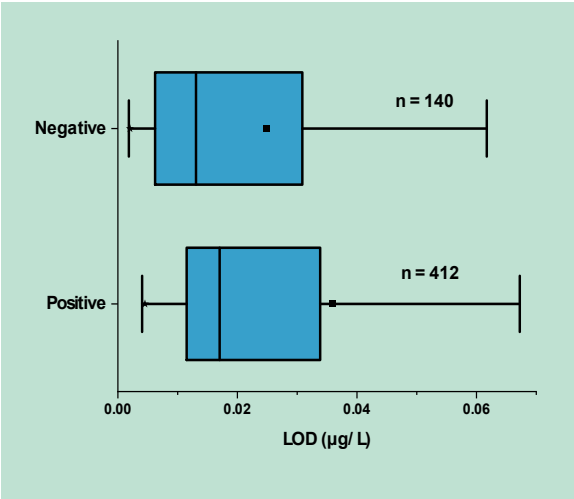

FIGURE 12: Box-plot of estimated detection limits in µg/L of nLC-ESI(-)-HRMS (top) and nLC-ESI(+)-HRMS (bottom).

3.2.2.2 IEC-HRMS and GC-HRMS

Estimated LOD's for the ion-exchange and gas chromatography platforms can be seen in Figure 13

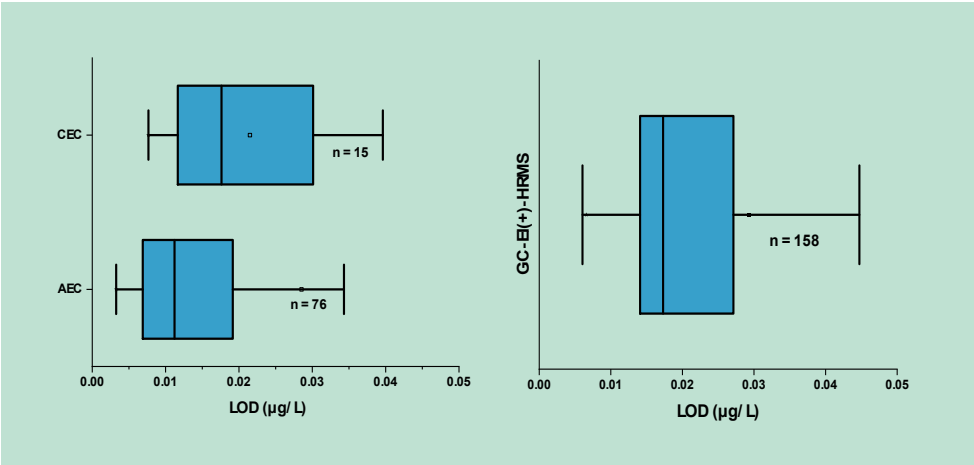

FIGURE 13: Left) Box-plot of estimated detection limits in µg/L of CEC-ESI(+)-HRMS (top) and AEC-ESI(-)-HRMS (bottom). **Right)** Box-plot of estimated detection limits in µg/L on GC-EI(+)-HRMS.

3.2.3 Sampling container material

Though no overall difference was seen between the usage of plastic and glass containers, the choice of container material could be important for the analysis of compounds presented in Table 3, where the choice of material resulted in a net-gain of more than 60 % on the recovery, or a large reduction in background.

TABLE 3: Compounds a large difference in recoveries based on container material. Their respective recommended container material is based on a recovery difference of >60% or a significant drop in background.

Compound	Recommended material
Benzoaminopurine	Plastic
Fluometuron	Plastic
Rimsulfuron-desulfon	Plastic
Thiofanox	Plastic
Desisopropy atrazine	Glass
Dibutyltin (DBT)	Glass
Imazalil	Glass
Metribuzin-diketo	Glass
Phenazone	Glass
Sulfamethoxazole	Glass

3.3 Non-targeted screening analysis

3.3.1 Detected compounds during quantitation

The NTS data processing pipelines of the un-spiked samples (DGU well number 191.265-2) indicated the presence of some of the 967 investigated xenobiotics. To fully confirm the presence of these compounds, samples should further be screened by a multi-step iterative approach [8] using suspect and inclusion lists containing these suspected compounds, followed by comparison of online and in-house spectral databases. The suspects were filtered, and excluded if their absolute peak areas were lower than 100,000, signal-to-noise ratios exceeded 3, recoveries exceeded 50%, and LODs were below 0.2 µg/L. Compounds succeeding in all criteria are listed in Table 4. Exact annotations are seen in Appendix 1.

TABLE 4: Number of suspects from DGU 191.265-2 unspiked quantitation samples on each respective platform.

Platform	Identified	Unique
nLC-ESI(+)-HRMS	115	109
nLC-ESI(-)-HRMS	9	4
CEC-ESI(+)-HRMS	1	1
AEC-ESI(-)-HRMS	7	5
GC-EI(+)-HRMS	6	6

3.3.2 Detected compounds in DGU 191.265-2

The groundwater sample DGU well number 191.265-2 was treated according to the applied optimal sample preparation and analysed across all platforms. Applying an optimised NTS data processing pipeline revealed more than 5,000 chemical substances in the groundwater sample. Further data filtering and curation resulted in 61 substances, from LC and IC platforms, that could be identified via MS2-spectral libraries (level 1 and 2) while another 160 substances were detected by GC could be annotated at level 2 and 3. The 61 level 1 and 2 compounds from LC and IC are shown in Table 5. A list of GC-EI(+) annotated compounds can be found in Appendix 3.

TABLE 5: Compounds annotated at either level 1 or level 2 [9], out of more than 5,000 substances, detected in a groundwater sample (DGU well number 191.265-2). Compounds identified by GC-EI(+)-HRMS are presented in Appendix 3.

Chemical	Platform	Level
Di-n-butyl phosphate	IEC(-)	1
Trifluoroacetic acid (TFA)	IEC(-)	1
2-Methyl-4-chlorophenoxyacetic acid (MCPA)	LC(-)	1
Butylparaben	LC(-)	1
Mecoprop	LC(-)	1
PFHxS	LC(-)	1
Salicylic acid	LC(-)	1
Diisobutyl phthalate	LC(+)	1
Epoxiconazole	LC(+)	1
Fenpropidin	LC(+)	1
Loratadine	LC(+)	1
Pyrimidifen	LC(+)	1
Spiroxamine	LC(+)	1
Tebutam	LC(+)	1
Tris(isobutyl) phosphate	LC(+)	1
(11ξ,12ξ)-4,7-Dihydroxy-12,13-epoxytrichothec-9-en-8-one	IEC(-)	2
2-Naphthalenesulfonic acid	IEC(-)	2
8-Hydroxy-4-(2-hydroxy-2-propanyl)-10-oxatricyclo[7.2.1.$0^{1,5}$]dodec-3-ene-8-carboxylic acid	IEC(-)	2
(±)9-HpODE	LC(-)	2
(15Z)-9,12,13-Trihydroxy-15-octadecenoic acid	LC(-)	2
(3aR,4R,5aS,6S,9aR,9bS)-4,6-Dihydroxy-5a,9-dimethyl-3-methylene-3a,4,5,5a,6,7,9a,9b-octahydronaphtho[1,2-b]furan-2(3H)-one	LC(-)	2
10-HDA	LC(-)	2
2,4-Dichlorophenoxypropionic acid	LC(-)	2
2,5-di-tert-Butylhydroquinone	LC(-)	2
3-tert-Butyladipic acid	LC(-)	2
4,4'-Dihydroxybenzophenone	LC(-)	2
4-Dodecylbenzenesulfonic acid	LC(-)	2
4-Hydroxybenzophenone	LC(-)	2
Azelaic acid	LC(-)	2
Cholic acid	LC(-)	2
Mono(2-ethylhexyl) phthalate (MEHP)	LC(-)	2
Monobutyl phthalate	LC(-)	2
Taurochenodeoxycholic acid	LC(-)	2
(-)-Camphor	LC(+)	2
(+/-)12(13)-DiHOME	LC(+)	2
(1R,5S,6S,8R)-8-Hydroxy-11-(hydroxymethyl)-1,5,11-trimethyltricyclo[6.2.1.$0^{2,6}$]undec-2-en-9-one	LC(+)	2
(1S,6S,8R,11R)-6-(Hydroxymethyl)-2,6,9-trimethyltricyclo[5.4.0.$0^{2,9}$]undecane-8,11-diol	LC(+)	2
(6E)-4,5,8-Trihydroxy-10-(2-hydroxypentyl)-3,3-dimethyl-3,4,5,8,9,10-hexahydro-2H-oxecin-2-one	LC(+)	2
(9Z,12E)-15,16-Dihydroxy-9,12-octadecadienoic acid	LC(+)	2
(E)-3,10-Dihydroxy-4,9-dimethyldodec-6-enedioic acid	LC(+)	2
12-Aminododecanoic acid	LC(+)	2
2,6-Di-tert-butyl-1,4-benzoquinone	LC(+)	2
2-[(7-methyl-2,3-dihydro-1H-inden-4-yl)oxy]pyridin-3-amine	LC(+)	2
2-{[3-(2,3,4,5,6-pentamethylphenyl)prop-2-ynyl]oxy}tetrahydro-2H-pyran	LC(+)	2
4-[(1E,3E)-1,3-Heptadien-1-yl]-3-(hydroxymethyl)-1,2-cyclohexanediol	LC(+)	2
4-Methylbenzotriazole	LC(+)	2
4-Methylumbelliferone hydrate	LC(+)	2

5-(Hydroxymethyl)-3-(1-hydroxy-4-methylpentyl)dihydro-2(3H)-furanone	LC(+)	2
5-Methylbenzotriazole	LC(+)	2
9S,13R-12-Oxophytodienoic acid	LC(+)	2
Cortisol	LC(+)	2
Cynaropicrin	LC(+)	2
Decanamide	LC(+)	2
Diacetoxyscirpenol	LC(+)	2
DOBU	LC(+)	2
Eicosatetraynoic acid	LC(+)	2
Meprednisone	LC(+)	2
N,N'-Dicyclohexylurea	LC(+)	2
Salvinorin B	LC(+)	2
Sertraline	LC(+)	2
Triphenylphosphine oxide	LC(+)	2
Valerophenone	LC(+)	2

3.4 Unresolved compounds

From the current metadata in Appendix 1, it was not possible to assign specific patterns in regards to unresolved compounds as they were spread out evenly across all physiochemical parameters. It is likely that the resolution of compounds depends stronger on the instrumental parameters such as solvent compositions, column materials, and ionisation energies. Structural metadata could help elucidate if a relationship between resolved compounds and their respective chemical groups and structures, though this is not pursued in this study.

A few observations could be taken from the dataset, namely that most volatile organic compounds and aromatics were unresolved - likely due to evaporation during sample preparation and storage. Also, most anticoagulants, large mass (>500 Da) flame retardants, and long-chained ethers, phthalates, and adipates were left unresolved. Many organometal(loid)s were also undetected, due to difficulties in both fragment identification and auto-processing algorithms. For these compounds, it is recommended to use, for example, combined LC-MS and ICP-MS techniques to resolve both chemical structures and metal(loid) composition. Also, some compounds of great relevance to current monitoring programmes such as 1,2,4-triazole and several atrazine transformation products remained unresolved. Most of these unresolved compounds are likely to be out-of-scope of the employed platforms except for the last mentioned monitored compounds, where additional effort could be put into further studies on different column materials [10].

3.5 Final remarks

Based on these findings we find that four of the five platforms, namely nLC-ESI(+)-HRMS, nLC-ESI(-)-HRMS, AEC-ESI(-)-HRMS, and GC-EI(+)-HRMS, are able to resolve 70 % of the investigated chemical space. Combining nLC-ESI(+)-HRMS and AEC-ESI(-)-HRMS allows for a rapid routine sampling and sample preparation protocols covering a wide chemical space. These two platforms combined are shown to detect up to 58% of the 967 xenobiotic compounds at levels of 0.1 µg/L in groundwater. Adding nLC-ESI(-)-HRMS as a third platform, increases the detection by up to 61 % and will include PFAS-like molecules.

The study was conducted on groundwater, and is likely to produce similar results for samples of drinking and rainwater. Likewise, the implementation of the presented NTS methodology on more complex water matrices such as coastal and/or fresh surface water, land leachate, and wastewater, should not pose immediate difficulties.

4. Conclusion

We investigated the coverage of a broad chemical space, modelled on detecting 967 environmental xenobiotics, on five NTS-platforms; nLC-ESI(+)-HRMS, nLC-ESI(-)-HRMS, AEC-ESI(-)-HRMS, CEC-ESI(+)-HRMS and GC-EI(+)-HRMS.

- In combination, the NTS platforms were able to detect 70% of the studied chemical space.
- Following a routine sample preparation of a 100-ng/L spiked groundwater sample, it was possible to recover up to 85% of the detectable chemical space.
- The two NTS-platforms nLC-ESI(+)-HRMS and AEC-ESI(-)-HRMS, in combination with the routine sample preparation, are complementary and able to detect up to 58% of the chemical space occurring at 100 ng/L in a groundwater sample — favouring hydrophilic and polar compounds.
- The three NTS-platforms nLC-ESI(+)-HRMS, nLC-ESI(-)-HRMS and AEC-ESI(-)-HRMS, in combination with the routine sample preparation, are complementary and able to detect up to 61% of the chemical space occurring at 100 ng/L in a groundwater sample — favouring both hydrophilic, polar, and acidic compounds.
- NTS platform pipelines and sample preparation methods can be optimised further to enhance the chemical space coverage and analyte recovery on both the LC and IEC platforms.

NTS can provide a snapshot of the current state of the groundwater quality. However, an annual and long-term program maybe provide basis for highly valuable and archived information, and the possibility for large-scale retrospective analysis.

5. Methods

Here follows a description of the applied methods and analytical platform used to complete the project activities.

5.1 Liquid chromatography high-resolution mass spectrometry

Liquid chromatographic separation was performed on a Dionex Ultimate 3000 NCS-3500RS Nano Proflow system (Thermo Scientific), configured with a trap and elute system enabling online sample enrichment and clean-up before ejection onto the analytical column. Ready samples were stored in glass vials and/or well-plates in a Dionex WPS-3000 TPL RS autosampler at 8°C. Dual-system gradients were used to successfully load analytes onto a C18 trap column (100 Å, C18, 0.3 mm x 5 mm, nanoViper, Thermo Scientific) followed by elution and ejection onto a nanoflow UHPLC column (PepMap RSLC, C18, 2 µm, 100 Å, 75 µm x 25 cm, Thermo Scientific). Both columns were kept at 40 °C. 5 µL sample was injected onto the trap column using full loop injection mode and a 5 µL sample loop with a flow of 30 µL/min loading solvent C (0.1 % formic acid in water) for 2 minutes. A valve switch at 2 and 18 minutes runtime connected and disconnected the trap column to the analytical column flow path respectively allowing the mobile phases to elute trapped. The flow rate of mobile phases was 300 nL/min. Chromatographic separation was achieved using a gradient beginning at 10 % mobile phase B (0.1 % formic acid in acetonitrile) and 90 % mobile phase A (0.1 % formic acid in water) kept for 2 minutes matching the time valve switch, at which point the gradient increased to 95 % for 15 minutes. This level was then maintained for 1 minute until the valve switch again disconnected the trap column from the analytical column. A level of 95 % B was kept for another 5 minutes through the analytical column. The conditions were restored to 10 % mobile phase B over 0.5 minutes followed by 6.5 minutes of equilibration time, leading to a total runtime of 30 minutes. After the valve switch at 18 minutes, the trap column was flushed with 30 µL/min 98 % loading solvent B (0.1 % formic acid in acetonitrile) and 2 % loading solvent C for 2 minutes followed by an equilibration step for 10 minutes using 100 % loading solvent C.

In between each injection the needle and fluidics were washed with 200 µL of 80 % acetonitrile and 0.1 % formic acid in water. The pump systems were rinsed every hour with a seal wash solution of 10 % methanol and 0.1 % formic acid in water. All solvents used were of UHPLC-MS grade.

The mass spectrometric analysis was performed on a high-resolution tandem mass spectrometer (Q Exactive HF, Thermo Scientific). Analytes were ionised by electrospray ionisation using an EASY-Spray ion source. The applied spray voltage was 1.50 kV during positive polarity and1.70 kV during negative polarity with a capillary temperature of 250 °C and an S-lens RF level of 50. No sheath, aux, and sweep gas was used.

HRMS acquisition was done in either full scan mode for quantification or data-dependent fragmentation (ddMS^2) mode for identification. Both the positive and negative polarity modes were used. Full scan acquisition was recorded using a resolution of 240K at m/z 200, an automatic gain control (AGC) target of 1e6, a maximum injection time of 200 ms, and a scan range of 70-1000 m/z. ddMS^2 acquisition was done using full scan settings with a resolution of 240K, AGC target of 1e6, maximum IT of 100 s, and scan range of 70-1000 m/z at m/z 200. ddMS^2 settings used a resolution of 15K, maximum IT of 50 s, an isolation window of 1.0 m/z, AGC target of 1e5, loop count of 10, and stepped collision energies of 15 and 50 NCE. The acquisition was performed with a dynamic exclusion of 20 s, minimum AGC target of 500, charge exclusion of

>3, and an apex trigger between 4-10 s. An estimated chromatic peak width (FWHM) was set to 8 s. An inclusion list of target ions of interest was used throughout every ddMS^2 acquisition run using a mass tolerance of 5 ppm and allowing the selection of other ions when idle. Sub-ppm mass accuracy was ensured by real time calibration of a lock mass of 371.10124 (polysiloxane from air) during positive polarity and 112.98563 (sodium formate cluster) during negative polarisation [11], [12]. Calibration of the mass spectrometer was performed with Pierce™ LTQ Velos ESI Positive and Negative Ion Calibration Solutions (Thermo-Fischer Scientific).

Instrumental performance was ensured by regular monitoring of an in-house laboratory quality control sample prepared from fetal bovine serum.

5.2 Ion exchange chromatography high-resolution mass spectrometry

State-of-the-art high-performance ion exchange chromatography was performed on a Dionex dual-pump ICS-6000 HPIC system (Thermo Scientific). Two ion exchange modes were used: Anion-exchange chromatography (AEC) and cation-exchange chromatography (CEC), each with separate hardware and chromatographic configurations.

5.2.1 Anion-exchange chromatography high-resolution mass spectrometry

A Dionex IonPac AS19-4 µm (2 x 250 mm) column was fitted with a Dionex AG19-4 µm (2 x 50 mm) Guard and connected to an ADRS 600 (2 mm) suppressor operated and 4.2 V (146 mA), and a conductivity detector cell. KOH was used as eluent, supplied by a Dionex KOH EGC 500 and regenerated by a Dionex CR-ATC 600.

During operation, the eluent was delivered at a flow rate of 0.45 mL/min at the following gradient settings: From 0 to 5 minutes 10 mM KOH, 5 to 11 minutes 10 to 60 mM KOH which was kept for 2 minutes followed by a sharp decrease back to 10 mM over 0.1 minute, which was kept for the remaining duration of the run with a total runtime of 20 minutes. Between 2 and 18 minutes, a timed valve switch enabled a steady flow of eluent to the MS. During this time, the eluent was mixed with a flow of 0.2 mL/min isopropanol, functioning as interface makeup solution, delivered by an external AXP pump.

The mass spectrometric analysis was performed on a high-resolution tandem mass spectrometer (Q Exactive HF, Thermo Scientific) in full scan and data-dependent acquisition mode. Analytes were ionised by electrospray ionisation using a HESI II source-probe. A spray voltage of 2.50 kV, capillary temperature of 380 °C, and S-lens RF level of 55 was used. Sheath, aux, and sweep gas flow rates were 32, 10, and 0 arbitrary units respectively, with an aux gas heater temperature of 350 °C.

Acquisition was done in either full scan mode (for quantification) or data-dependent fragmentation (ddMS^2) mode (for identification), both using negative polarity. For full scan acquisition, a resolution of 240K at m/z 200, AGC target of 1e6, maximum IT of 250 ms, and a scan range of 60-900 m/z at m/z 200 was used. For ddMS^2 acquisition, the full scan settings used a resolution of 240K, AGC target of 1e6, maximum IT of 100 s, and scan range of 60-900 m/. The following ddMS^2 settings used a resolution of 15K, maximum IT of 100 s, isolation windows of 0.7 m/z, AGC target of 1e5, loop count of 10, and stepped collision energies of 15 and 50 NCE. The acquisition was performed with a dynamic exclusion of 5 s, minimum AGC target of 1e3, charge exclusion of >3, and no apex trigger. An estimated chromatic peak width (FWHM) was set to 6 s. An inclusion list of target ions of interest was used throughout every ddMS^2 acquisition run using a mass tolerance of 5 ppm and allowing the selection of other ions when idle. Sub-ppm mass accuracy was ensured by real time calibration of a lock mass of 112.98563 (sodium formate cluster) [12]. Calibration of the mass spectrometer was performed with Pierce™ LTQ Velos ESI Negative Ion Calibration Solution (Thermo-Fischer Scientific).

5.2.2 Cation-exchange chromatography high-resolution mass spectrometry

A Dionex IonPac CS17 (2 x 250 mm) column was fitted with a Dionex CG17 (2 x 50 mm) Guard and connected to a CDRS 600 (2 mm) suppressor and a conductivity detector. Methansulfonic acid (MSA) was used as eluent and supplied by a Dionex MSA EGC 500, regenerated by a Dionex CR-CTC 600.

During operation, the eluent was delivered at a flow rate of 0.4 mL/min at the following gradient settings: From 0 to 5 minutes 6 mM MSA, 5 to 11 minutes 6 to 60 mM MSA which was kept for 2 minutes followed by a sharp decrease back to 6 mM MSA over 0.1 minute, which was kept for the remaining duration of the run with a total runtime of 20 minutes. Between 2 and 18 minutes, a timed valve switch enabled a steady flow of eluent to the MS. During this time, the eluent was mixed with a flow of 0.2 mL/min acetonitrile, functioning as interface makeup solution, delivered by an external AXP pump.

The mass spectrometric analysis was performed on a high-resolution tandem mass spectrometer (Q Exactive HF, Thermo Scientific). Analytes were ionised by electrospray ionisation using a HESI II source-probe. A spray voltage of 4 kV, capillary temperature of 425 °C, and S-lens RF level of 50 was used. Sheath, aux, and sweep gas flow rates were 40, 10, and 1 arbitrary units respectively, with an aux gas heater temperature of 260 °C.

Acquisition was done in either full scan mode (for quantification) or data-dependent fragmentation ($ddMS^2$) mode (for identification), both using positive polarity. For full scan acquisition, a resolution of 240K at m/z 200, AGC target of 1e6, maximum IT of 100 ms, and a scan range of 60-900 m/z was used. For $ddMS^2$ acquisition, the full scan settings used a resolution of 240K at m/z 200, AGC target of 1e6, maximum IT of 100 s, and scan range of 60-900 m/z at m/z 200. The following $ddMS^2$ settings used a resolution of 15K at m/z 200, maximum IT of 100 s, isolation windows of 0.7 m/z, AGC target of 1e5, loop count of 10, and stepped collision energies of 15 and 50 NCE. The acquisition was performed with a dynamic exclusion of 10 s, minimum AGC target of 1e3, charge exclusion of >3, and no apex trigger. An estimated chromatic peak width (FWHM) was set to 14 s. An inclusion list of target ions of interest was used throughout every $ddMS^2$ acquisition run using a mass tolerance of 5 ppm and allowing the selection of other ions when idle Sub-ppm mass accuracy was ensured by real time calibration of a lock mass of 144.98215 (copper-acetonitrile cluster) [12]. Calibration of the mass spectrometer was performed with Pierce™ LTQ Velos ESI Positive Ion Calibration Solution (Thermo-Fischer Scientific).

Both ion-exchange suppression modes were run using external UltraPure water supply at 0.45 mL/min, delivered to each respective suppressor by an external AXP pump. To prevent unwanted system damage, a script was implemented to shut down the system if a conductivity of >100 µS was measured for longer than 30 seconds.

Ready samples were stored in polypropylene vials and/or polypropylene well-plates in a Dionex AS-AP autosampler at 8 °C, where they were transported to a 25 µL sample loop via a 90 µL transfer line. Sample injection volumes were 12 µL using limited_solvent injection mode. The temperatures of the detector cell, columns, and suppressors were 45 °C, 40 °C, and 35 °C respectively.

Instrumental performance was ensured by regular monitoring of in-house laboratory quality control samples containing small cations and anions (Thermo Scientific), and a sample containing a mixture of cationic and anionic species supplied by NEOCHEMA GmbH.

UltraPure water (18.2 MΩ cm) was provided by a Dionex IC Pure Water Purification System (Thermo Scientific).

5.3 Gas chromatography high-resolution mass spectrometry

State-of-the-art Orbitrap based GC-HRMS was achieved using an Exactive GC system equipped with a TriPlus autosampler (Thermo) and a TraceGOLD TG-5MS analytical column (60 m, 0.25 µm, 0.25 mm, 5% phenyl - 95% dimethyl polysiloxane phase, ThermoFisher Scientific) installed in a TRACE 1310 GC (ThermoFisher Scientific). One-microliter sample extract was injected sandwiched with air using a split-splitless mode at 280 °C and 70 mL/min split flow after 60 sec. The column was operated with high purity helium at 1.00 mL/min and a temperature program; initial 60 °C with 2 min hold and ramped (5 °C/min) to 240 °C and further (10 °C/min) to 300 °C with a final holding time of 16 min. Analytes were transferred using a MS-transferline at 280 °C and ionized using electron impact ionisation (EI) at 70 eV with a 12 minutes filament delay. The Orbitrap HRMS system was operated in full scan mode (m/z 50 to 750) at a 60,000 resolution in centroid mode and an automatic gain-control target of 1e6 ions. The Q Exactive HRMS system was tuned and calibrated on a daily basis using FC43.

5.4 Sampling

Twelve 2 L grab samples of groundwater were collected from DGU well number 191.265-2 in November 2020. Six samples were collected in glass containers and six in HDPE containers. The samples were stored at 4 °C immediately after collection and until extraction. In addition to this, two CLAM-samples were collected on-site: one HLB and one C18. Each disk was loaded with approximately 5 L of sample by connecting them directly to the pump-outlet with an approximate flow of 500 mL/min for ten minutes. Two pre-conditioned CLAM-disks (one HLB and one C18) were brought to the field to serve as field-blanks. For further description of the application of CLAM-disks, refer to our previous work described in [1].

5.5 Sample preparation

5.5.1 Preparation of standards

Analytically pure standards were purchased as either mixtures or neat compounds. Suppliers were Thermo-Fisher Scientific, Sigma-Aldrich, Restek, HPC Standards GmbH, NEOCHEMA GmbH, LGC Standards, and Alfa Aesar. Around 500 standards were already present in our laboratory, purchased in the period of 2018 to 2020. All 967 standards were stored at -20 °C.

To simplify storage, sample preparation, and data processing, the 967 analytical standards were combined into 29 different stock solutions of 10 µg/mL (where possible) in 5 % methanol or acetonitrile (where possible). As most of the 967 compounds were delivered in 10 or 100 µg/mL pre-mixed stock solutions, these were mixed according to matching solvent composition, either water, methanol, acetonitrile, methyl tert-butyl ether, dichloromethane, cyclohexane, or isooctane. Some of these solutions were additionally stabilised using either small amounts of H_3PO_4, formic acid, DMSO, or acetone. For the remaining few compounds that were purchased as either solid or liquid neat standards, these were mixed according to their presumed stability and solubility in either water, acetonitrile, or tetrahydrofuran. Using these stock solutions, 29 standard mixtures were prepared containing 1 µg/mL analyte in either 5 % methanol or 5 % acetonitrile (and for some mixtures, small amounts (<0.1-1 %) of the aforementioned alternative solvents and stabilisers). These standard mixtures were aimed to be compatible for LC (and IC) analysis, where high organic content can lead to loss of analytes in the trap and elute system described in section 5.1. In addition to this, the standard mixtures were prepared as such to reduce the amount of isobaric overlap within each standard – giving an advantage during later post-processing steps.

To make the standard compatible for GC-HRMS analysis and to avoid injecting water unto the GC column, GC specific standard mixtures were made by liquid-liquid extracting analytes from 100 µL of each standard mixture into 100 µL isooctane.

Exact concentration, solvent composition, and standard mixture number can be seen for each respective compound in Appendix 1.

5.5.2 Solid-phase extraction

The groundwater samples were enriched by tandem solid-phase extraction (SPE) using a mixed-anion exchange (MAX) (Oasis MAX 6 cc 500 mg) cartridge and a graphitised non-porous carbon (Supelclean ENVI-Carb 500 mg) cartridge. This was anticipated to cover a large chemical space due to the hydrophilic-lipophilic balanced (HLB) skeleton and positively charged functional groups of the MAX cartridge in addition to the lipophilic affinity of the carbon-based cartridge. The MAX cartridge was conditioned and equilibrated using 5 mL methanol followed by 5 mL water. The ENVI-carb cartridge was conditioned and equilibrated using 5 mL acetone, then 5 mL methanol, and lastly 5 mL water. Each 2 L sample was loaded unto the cartridges in tandem. The cartridges were then washed using a 5 % NH_4OH solution in water. The two cartridges were then eluted separately, with the MAX cartridge being eluted with 5 mL methanol followed by 5 mL 0.1 % formic acid solution in methanol, and the ENVI-carb cartridge being eluted by 2 x 5 mL acetone [13]–[15]. Eluates from both cartridges were then combined into a mixture totalling 20 mL, which could then be spiked with 200 ng analyte (20 µL of 10 µg/mL stock solution) to produce post-spiked samples (described in section 5.5.4). The mixture was then split into two 10 mL fractions, one to be used for GC analysis, the other for LC and IC analysis. The LC and IC fraction was evaporated to dryness under N_2, reconstituted to 1 mL 5 % methanol and stored in Eppendorf tubes at -20 °C until analysis. The GC fraction was evaporated to near-dryness under N_2 using 300 µL isooctane as keeper, reconstituted to 1 mL isooctane, and stored in Eppendorf tubes at -20 °C until analysis.

The elution scheme for CLAM samples have been described in our earlier work [1], with the addition of an elution scheme for GC analysis, where samples are reconstituted in isooctane rather than 5 % methanol.

5.5.3 Derivatisation of samples for GC-HRMS

The effect of derivatisation was investigated on a subset of GC samples. 50 µL sample was evaporated to dryness using a SpeedVac (Savant, Thermo Scientific) and reconstituted in a solution of 90 % N,O-bis(trimethylsilyl)trifluoroacetamide and 10 % trimethylchlorosilane. This mix was then vortexed and heated to 80 °C for 60 minutes before use in GC analysis.

5.5.4 Spike-recovery

A total of 12 samples were analysed, 6 in glass containers, and 6 in plastic (HDPE) containers. Two of each sample type was spiked before undergoing sample preparation (pre-spiked), and two after sample preparation (post-spiked). Two of each sample type was kept unspiked. Samples were spiked by addition of 20 µL each of 18 individual 10 µg/mL (200 ng) standard to a final concentration of 0.1 µg/L for most compounds (see Appendix 1 for exact spiking concentrations.) Pre-spiked samples were spiked 12 hours before extraction and stored at 4 °C.

Recoveries were calculated from the average peak areas of pre, post, and unspiked samples according to

$$\mathrm{rec\%} = \frac{\overline{\mathrm{Pre}} - \overline{\mathrm{no}}}{\overline{\mathrm{Post}} - \overline{\mathrm{no}}} \cdot 100\%$$

where rec% is the is recovery in percent, $\overline{\mathrm{Pre}}$ is the average peak area in the pre-spiked sample, $\overline{\mathrm{Post}}$ is the average peak area in the post-spiked sample, and $\overline{\mathrm{no}}$ is the average peak area in the unspiked sample. The relative standard deviation of the recoveries were calculated by

$$\frac{\sigma}{rec\%} = \sqrt{\left(\frac{\sqrt{\sigma_{Pre}^2 + \sigma_{no}^2}}{\overline{Pre} - \overline{no}}\right)^2 + \left(\frac{\sqrt{\sigma_{post}^2 + \sigma_{no}^2}}{\overline{Post} - \overline{no}}\right)^2}$$

where σ is the standard deviations for pre, post, and unspiked samples respectively [16].

5.5.5 Detection limits by standard addition

To estimate detection limits, the method for non-weighted linear calibration described in ISO 11843-2 was applied [17]–[19]. Five calibration standards were prepared in sample matrix by combining the four post-spiked samples into a single pool and diluting according to the desired dilution scheme with a single pool of the four unspiked samples. The calibration reference states were 0 µg/mL, 0.01 µg/mL, 0.02 µg/mL, 0.05 µg/mL, and 0.10 µg/mL respectively for most compounds - corresponding to µg/L levels with an enrichment factor of 1000. Each reference state was prepared in triplicate. Some compounds could not be added in the aforementioned concentrations due to availability limitations. Instead, the highest calibration point (spike concentration) following an identical dilution scheme can be found for each compound in Appendix 1.

5.6 Post-processing

Due to the size of the resulting data sets, several pipelines were employed in order to reduce data complexity and processing time. To simulate a real non-targeted acquisition workflow, the pipelines were built around already established acquisition parameters and signal processing software.

5.6.1 Identification of standards

Each of the 29 standard mixtures were recorded in ddMS^2-mode on each analytical platform and processed using Compound Discoverer 3.2. Raw data was imported and processed using a fully automated NTS workflow with high selection criteria for each respective platform, producing a list of identified compounds with corresponding retention times for each platform. To further increase the hit-rate of compounds, data was reacquired using inclusion lists of the remaining unidentified compounds, reprocessed using similar settings, and lastly reprocessed using selection criteria lower by an order of magnitude. Retention times and parent ions were assigned to each identified compound and used as selection criteria in proceeding data processing. Several of the identified compounds and retention times were confirmed by literature. References are mentioned in the respective comments fields in Deliverance 1 (Appendix 1) [20]–[34].

5.6.2 Compound Discoverer 3.2

The Compound Discoverer 3.2 software was employed for identification of compounds in the standard mixtures and for non-targeted analysis of the water samples collected at DGU well number 191.265-2.

5.6.2.1 GC deconvolution

The ability to deconvolute GC data came with the release of Compound Discoverer 3.2. This enabled fast NTS processing with the ability of using NIST reference libraries. Workflow customisation was, however, very limited due its recent release date. For this reason, the identification of compounds for GC-EI(+)-HRMS was combined with a screening method in TraceFinder 4.1. Selection criteria was based on mass list search, S/N>100, and having an isotopic pattern score greater than 70 %.

5.6.3 mzVault 2.3 library generation

Identified compounds with fitting MS^2-data were added to an mzVault v2.3 database as entires for in-house spectral and retention time references for level one annotation confidence [9]. Libraries were generated by performing retention time based MS^2 search using highest intensity

matching, recalibration, threshold application, minimum intensity of 10,000, and a mass accuracy of 2 ppm.

5.6.4 Quantitation in TraceFinder EFS 4.1

Peak areas were obtained from an EFS quantitation workflow using the software TraceFinder v4.1. Retention-time based, standard ICIS peak detection settings were employed on all data sets using a compound database of all target masses of interest. Following automatic peak integration, each peak was manually reviewed to ensure high-quality and consistent data across the thousands of integrated peaks.

Data had been recorded by full-scan mode. A pooled QC sample was recorded in ddMS^2 and used to verify peaks with in case of matrix induced interference or retention time shifts

Calculated peak areas were exported as combined Excel-sheets used for direct import into a custom MATLAB script.

5.6.5 Automated pipeline in MATLAB

Estimated detection limits were calculated using a custom MATLAB-script following the computation method for the minimum detectable value described in ISO 11843-2 [19]:

$$x_d = 2t_{0.95}\frac{\hat{\sigma}}{\hat{b}}\sqrt{\frac{1}{K}+\frac{1}{J\cdot I}+\frac{\bar{x}^2}{s_{xx}}}$$

$\hat{\sigma}$ is the estimated residual standard deviation, $\hat{b}$ the estimated slope of the regression, K the number of preparations of the actual state (K = 1), J the number preparations for each reference state (J = 3), I the number of reference states (I = 5), $\bar{x}$ the average value of the calibration reference states ($\bar{x}$ = 36 µg/L), s_{xx} the sum of squared deviations of the net state variable values for the reference states from the average, $t_{0.95}$ students *t*-value for $\nu = J\cdot I-2$ degrees of freedom (ν = 13) at a significance level of 5 % (t = 1.77), and x_d the estimated detection limit in µg/L. The MATLAB script was verified by successfully running C.1 Example 1 found in [19]. Estimated detection limits for compounds following different calibration scheme concentrations were done by post-correction according to their actual calibration scheme concentrations.

An additional automated pipeline for the calculation of recoveries and their respective variances was also written in MATLAB.

5.6.6 Dynamic concentration ranges

Performing automatic (un-supervised) linear regression led to several complications in regards to data quality. Generally speaking, the data looked good for both the gas and ion-exchange platforms. In the case of LC-HRMS, however, the data posed several challenges in the form of exceeding working ranges and heteroscedastic behavior - both of which negatively affects the LOD estimation. An example of this is shown in Figure 14, where calibration curves are shown for three compounds: 2,6-dichlorprop, benalaxyl, and buturon, autoprocessed from nLC-ESI(+)-HRMS data. These three curves perfectly exemplify the challenges in regards to data quality. The calibration curve for benalaxyl shows an acceptable linearity within the dynamic range and an estimated LOD of < 0.02 µg/L. The curve for 2,6-dichlorprop exhibits heteroscedastic behavior, where the standard deviation increases proportionally with the concentration. This - in principle - disqualifies the dataset from being processed according to a non-weighted approach, as reference states in this case are required to have (near-)constant standard deviations. Though this should be addressed in a revised MATLAB workflow, it was ignored for the time being as only 36 compounds showcased this behavior - and only seemed to affect the results in the LC-HRMS platform. The estimated detection limits in these cases would be grossly overestimated. Lastly, an example of data exceeding the working range is seen with the 5-point calibration curve

for buturon, where the regression residuals became smaller by instead performing 4-point regression to stay closer within the linear working range.

In addition to these factors, LODs were additionally overestimated for compounds exceeding spiking concentrations of 1.0 µg/mL due to the nature of the model. This would result in unnaturally large LOD values for certain compounds, seen for example for haloacetic acids, fragrances, and MCPA and mecoprop, that were all spiked at higher levels due to concentration limits of the commercially supplied standards.

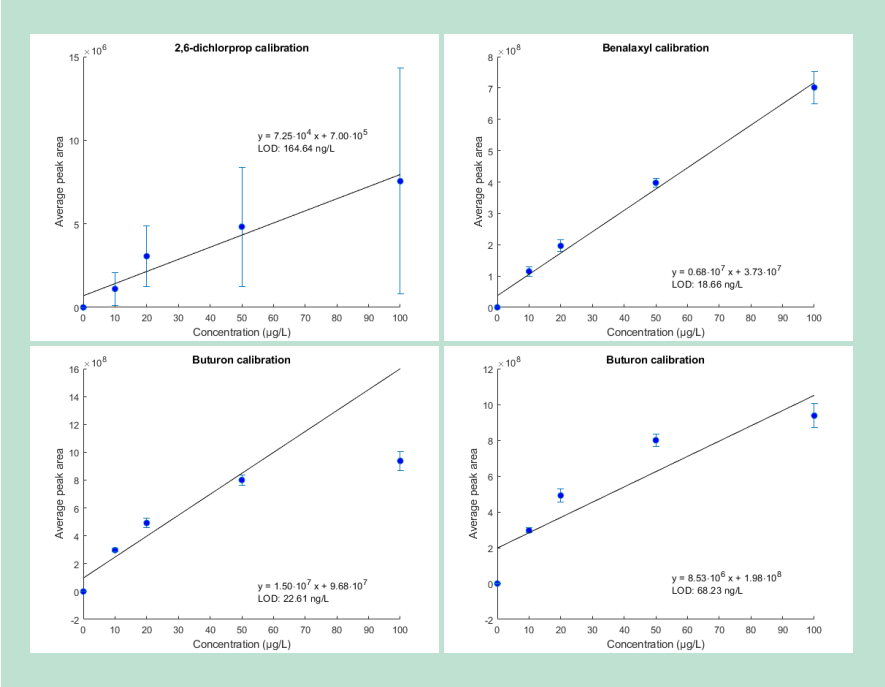

FIGURE 14: Calibration curves auto-processed by un-supervised MATLAB-workflow. 2,6-dichlorprop (top left) showcasing heteroscedasticity, benalaxyl (top right) showcasing 'ideal' behaviour, and buturon showing how an improvement in linearity can be achieved by reducing the number of reference states from five (bottom left) to four (bottom right) in the calibration curve.

6. References

[1] M. Hansen *et al.*, “Danish EPA pesticide and biocide research programme, project: HITLIST, grant no. MST-667-00207, project period: June 2018 to December 2019.,” 2020.

[2] T. K. O. Gravert, J. Vuaille, J. Magid, and M. Hansen, “Non-target analysis of organic waste amended agricultural soils: Characterisation of added organic pollution.,” Under review, 2021.

[3] N. L. Ma *et al.*, “Body mass, mercury exposure, biochemistry and untargeted metabolomics of incubating common eiders (Somateria mollissima) in three Baltic colonies,” *Environ. Int.*, 2020, doi: 10.1016/j.envint.2020.105866.

[4] J. R. Sobus *et al.*, “Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA,” *Journal of Exposure Science and Environmental Epidemiology*. 2018, doi: 10.1038/s41370-017-0012-y.

[5] E. Szöcs, T. Stirling, E. R. Scott, A. Scharmüller, and R. B. Schäfer, “Webchem: An R package to retrieve chemical information from the web,” *J. Stat. Softw.*, 2020, doi: 10.18637/jss.v093.i13.

[6] D. Gadaleta, A. Lombardo, C. Toma, and E. Benfenati, “A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications,” *J. Cheminform.*, vol. 10, no. 1, pp. 1–13, 2018, doi: 10.1186/s13321-018-0315-6.

[7] D. Zahn, I. J. Neuwald, and T. P. Knepper, “Analysis of mobile chemicals in the aquatic environment—current capabilities, limitations and future perspectives,” *Analytical and Bioanalytical Chemistry*. 2020, doi: 10.1007/s00216-020-02520-z.

[8] J. P. Koelmel *et al.*, “Acquisition With Automated Exclusion List Generation,” *J Am Soc Mass Spectrom.*, vol. 28, no. 5, pp. 908–917, 2018, doi: 10.1007/s13361-017-1608-0.Expanding.

[9] E. L. Schymanski *et al.*, “Identifying small molecules via high resolution mass spectrometry: Communicating confidence,” *Environ. Sci. Technol.*, vol. 48, no. 4, pp. 2097–2098, 2014, doi: 10.1021/es5002105.

[10] B. Buszewski and S. Noga, “Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique,” *Analytical and Bioanalytical Chemistry*. 2012, doi: 10.1007/s00216-011-5308-5.

[11] A. Schlosser and R. Volkmer-Engert, “Volatile polydimethylcyclosiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry,” *J. Mass Spectrom.*, vol. 38, no. 5, pp. 523–525, 2003, doi: 10.1002/jms.465.

[12] J. V. Olsen *et al.*, “Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap,” *Mol. Cell. Proteomics*, 2005, doi: 10.1074/mcp.T500030-MCP200.

[13] J. C. Arsenault, *Beginner’s Guide to SPE: Solid-Phase Extraction*, 1st ed. Waters, 2012.

[14] C. Huang *et al.*, “Preparation of a reversed-phase/anion-exchange mixed-mode spherical sorbent by Pickering emulsion polymerization for highly selective solid-phase extraction of acidic pharmaceuticals from wastewater,” *J. Chromatogr. A*, 2017, doi: 10.1016/j.chroma.2017.09.021.

[15] E. Pérez-Carrera *et al.*, “Multiresidue method for the determination of 32 human and veterinary pharmaceuticals in soil and sediment by pressurized-liquid extraction and LC-MS/MS,” 2010, doi: 10.1007/s00216-010-3862-x.

[16] D. C. Harris, *Quantitative Chemical Analysis*, vol. 8th. 2010.

[17] T. Wenzl, J. Haedrich, A. Schaechtele, P. Robouch, and J. Stroka, *Guidance Document on the Estimation of LOD and LOQ for Measurements in the Field of Contaminants in Feed and Food. EUR 28099 EN*. 2016.

[18] Eurachem, *The Fitness for Purpose of Analytical Methods*. 1998.

[19] B. S. Iso, “Capability of detection - Part 2: Methodology in the linear calibration case British Standard - ISO 11843-2:2000,” vol. 3, no. October, pp. 1–34, 2009.

[20] Waters, *Cosmetics and Personal Care Application Notebook*. 2016.

[21] BASF, “Analytical Method No. 01410 (L0257/01) for the determination of residues of the geometric isomers of dimethomorph, E-dimethomorph and Z-dimethomorph, in drinking (tap) water and surface (pond) water by LC-MS/MS.,” 2016.

[22] J. Liu, K. Hara, S. Kashimura, T. Hamanaka, S. Tomojiri, and K. Tanaka, “Gas chromatographic-mass spectrometric analysis of dichlorobenzene isomers in human blood with headspace solid-phase microextraction,” *J. Chromatogr. B Biomed. Sci. Appl.*, 1999, doi: 10.1016/S0378-4347(99)00226-1.

[23] C. Anyakora, A. Ogbeche, P. Palmer, H. Coker, G. Ukpo, and C. Ogah, “GC/MS analysis of polynuclear aromatic hydrocarbons in sediment samples from the Niger Delta region,” *Chemosphere*, vol. 60, no. 7, pp. 990–997, 2005, doi: 10.1016/j.chemosphere.2004.12.073.

[24] Restek, “GC Multiresidue Pesticide Standard #2-OCP on Rxi®-5ms by GC-MS.” https://www.restek.com/images/cgram/gc_fs0601.pdf (accessed Feb. 17, 2021).

[25] L. Webster, R. Patrick, P. Bersuder, M. Kotterman, M. Haarich, and K. Vorkamp, “Determination of polychlortinated biphenyls (PCBs) in sediment and biota,” *ICES Tech. Mar. Environ. Sci.*, vol. 53, p. 18, 2013.

[26] L. Wolska, P. Konieczka, A. Jastrzebska, and J. Namieśnik, “Analytical procedure for the determination of chlorobenzenes in sediments,” *J. Chromatogr. Sci.*, 2003, doi: 10.1093/chromsci/41.2.53.

[27] C. D. S. Tomlin, “The Pesticide Manual,” in *The Pesticide Manual*, 13th ed., Surrey UK: British Crop Protection Council, 2003.

[28] R. Pandey, P. Chandra, M. Srivastva, K. R. Arya, P. K. Shukla, and B. Kumar, “A rapid analytical method for characterization and simultaneous quantitative determination of phytoconstituents in Piper betle landraces using UPLC-ESI-MS/MS,” *Anal. Methods*, 2014, doi: 10.1039/c4ay00975d.

[29] Q. Zhang, M. Wang, M. Tian, M. Wang, and H. Shi, “Simultaneous enantioselective determination of triazole fungicide flutriafol in vegetables, fruits, wheat, soil, and water by reversed-phase high-performance liquid chromatography,” *J. Agric. Food Chem.*, 2014, doi: 10.1021/jf405689n.

[30] Y. Suzuki, T. Kaneko, and K. Saito, “The internal standard diquat-d 4 causes errors in diquat analysis by LC–MS/MS,” *Forensic Toxicol.*, 2018, doi: 10.1007/s11419-018-0423-z.

[31] F. Sacher, B. Raue, and H. J. Brauch, “Analysis of iodinated X-ray contrast agents in water samples by ion chromatography and inductively-coupled plasma mass spectrometry,” 2005, doi: 10.1016/j.chroma.2005.01.031.

[32] F. Buiarelli *et al.*, “Hydrophilic Interaction Liquid Chromatography-Tandem Mass Spectrometry Analysis of Fosetyl-Aluminum in Airborne Particulate Matter,” *J. Anal. Methods Chem.*, 2018, doi: 10.1155/2018/8792085.

[33] R. Sandín-España, J. J. González-Blázquez, J. O. Magrans, and J. M. García-Baudín, “Determination of herbicide tepraloxydim and main metabolites in drinking water by solid-phase extraction and liquid chromatography with UV detection,” *Chromatographia*, 2002, doi: 10.1007/BF02491782.

[34] J. Cole, R. Law, and C. Cojocariu, “Detection and Quantification of Fragrance Allergens in Complex Matrices Using GC Orbitrap MS Technology,” 2019.

[35] Martin, “chemtrans,” *MATLAB Central File Exchange*, 2021. https://www.mathworks.com/ matlabcen-tral/fileexchange/68343-chemtrans (accessed Jul. 19, 2020).

Appendix 1. Deliverance 1A

Appendix 1 is Deliverance 1A
Here is a link to Appendix 1. The Appendix can also be found through mst.dk/service/publikationer.

Appendix 2. Deliverance 1B

Initially, an 'interest list' of 1772 xenobiotic substances and endogenous metabolites was generated. This list was later revised to a 967-substance suspect list – only including xenobiotics (MFS). This appendix describes how this list was established. The total number of substances presented in this appendix are based on several steering group meetings and discussions in 2020 and were left primarily unchanged to match the numbers in a project deliverable (August 2020) and for transparency reasons.

A list of over 1000 xenobiotic compounds was generated for the purpose of validating non-targeted screening (NTS) methodology. Targeting more than 1000 xenobiotic compounds the project will investigate the chemical space covered in three different high-resolution mass spectrometric NTS platforms (viz. liquid, gas, and ion chromatography). This chapter describes the selection criteria for the selected compounds. The complete substance list can be found in Appendix 1 and are available as an Excel file.
In order to build a compound list covering the desired chemical space, several criteria were followed:

- The primary objective was to include as many compounds used in water monitoring programmes as possible. This was done by consultation with the Danish Environmental Protection Agency and by referring to several national and international directives, seen in the following sections. Emphasis was additionally placed on selecting compounds that are expected to be found in the aquatic environment, while simultaneously covering a large range of both compound classes and the chemical functional groups.
- As the selected compounds are provided from commercial sources, a requirement was that these substances are commercially available and within a reasonable budget framework.

To further extend the chemical space of compounds analysed in the project, an additional 1339 in-house standards were added to the compound list.

The chemical space
After extensive research, the original list ended up consisting of 1772 different xenobiotic compounds relevant to water monitoring. Compound descriptions are displayed in Appendix 1 containing Classification, Name, Formula, CAS and Directive for each compound.

Information of compound classifications were obtained from commercial suppliers, Chemspider (Royal Society of Chemistry) and internet searching. The included chemical classifications are listed in Table A1, and in illustrated in Figure A1.

Compound names are written as they are either represented by their Directives or as defined by either suppliers or Chemspider. Chemical formulas are obtained directly from Directives or suppliers, or datamined from the CACTUS NCI/CADD-database using in-house MATLAB scripting [35]. CAS-numbers are similarly obtained from either Directives or suppliers, or datamined from the ChemIDplus (SRC, Inc.) database using the webchem R-scripting package [5]. For compounds found in Directives, only one Directive is mentioned for each compound - even if it has multiple entries. Compounds assigned as 'EML-AU' indicates compounds we have selected out of interest from previous studies or because these compounds were purchased as part of commercial mixtures.

Table A1: Chemical classifications for the 1772 selected compounds.

Classifications	Amount	Classifications	Amount
Alkylphenols	2	Lipophilic metabolites	114
Anionic detergents	1	Nitrophenols	9
Antibiotics	7	Opiates	3
Anticoagulants	10	Organic acid metabolites	94
Aromatics:	12	Organotin compounds	4
Artificial sweeteners	7	Other organometal(loid)s	7
Bile acids	33	PAHs	26
Carnitines	32	Parabens	5
CFCs	1	PCBs	7
Cholesteryl esters	6	Pesti-, herbi-, fungicides	587
Drugs	2	PFAS	23
Endocrine disruptors	52	Pharmaceuticals	98
Fatty acid metabolites	96	Phenols	6
Flame retardants	4	Plasticisers	8
Fragrances	39	P-triesters	3
Haloacetic acids	9	Steroid hormones	11
Halogenated aliphatic hydrocarbons	12	Sterols	23
Hydrophilic metabolites	400	UV filters	1
Industrial chemicals	2	VOCs	8
Insect repellents	3	X-ray contrast media	5

CFCs, chlorofluorocarbons; PAHs, polycyclic aromatic hydrocarbon; PCBs, polychlorinated biphenyls; PFAS, per- and polyfluoroalkylated substances; VOC, volatile organic carbons.

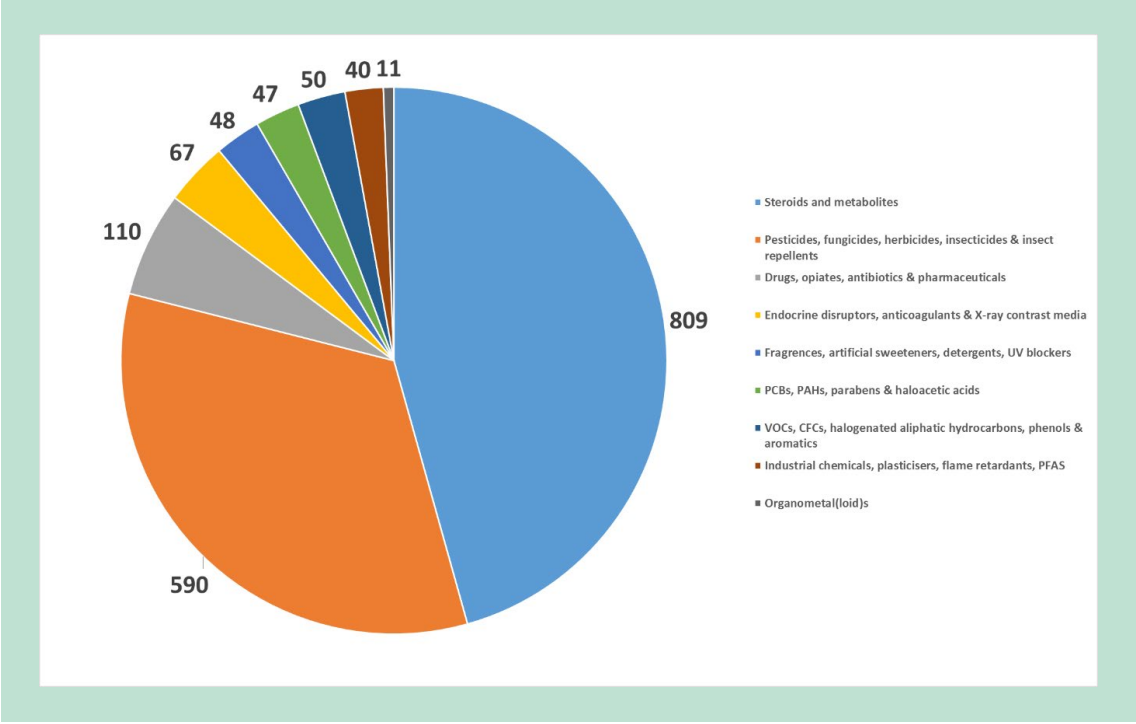

Figure A1: Diagram displaying compound classes of the 1772 investigated substances.

It is anticipated that the compound list will be further expanded with the addition of additional chemical identifiers (InChIKey and SMILES), structures (mol-file/SDF), physiochemical properties (logP, pKa, water solubility, Henry's constant, vapour pressure, boiling point), and toxicological properties (e.g. LC_{50}). This will be done using the aforementioned scripting workflows. In addition to this, confirmed analytical platform(s) and detection limits will be added once experimental information is obtained. For relevant compounds, the current legislated detection limits will be mentioned as well.

Of the 1772 compounds, 433 were purchased from commercial sources specifically for this project. Efforts was made into deciding exactly which compounds to include in the study, as the inevitably exclusion of some compounds of interest disallowed a full representation of the Directives and other desired compounds.

Due to budgetary reasons, certain legislated compounds were omitted from these investigations. The excluded compounds were either not commercially available, restricted, too expensive, highly reactive and otherwise unstable, or chemically similar to another chosen compound. Exact descriptions are written for each excluded compound in Appendix 1.

Certain legislative directives refer to whole compound groups rather than individual targets. As for example seen for alkylbenzenesulphonates and nonylphenols. Representative compounds were selected for each of these nonspecific compound groups, allowing for a few compounds to represent a whole group of chemically similar compounds. This would significantly reduce the costs while losing little information in regards to the identification of these compound groups.

Additionally, legislated anions and oxyhalides were not included under the assumption that anions are well resolved in the used ion-chromatographic analysis platform. An anion standard solution is available in-house and will be used to confirm the identification capability of several anionic species.

The employed analytical platforms were also not suitable for trace element analysis. Aside from a select few commercially available organometallic compounds, the legislated trace elements were not pursued in this project.

As the compound list contains over a thousand entities, acquisition of identifiers and properties were largely done by employing in-house MATLAB and R-scripting workflows to access chemical databse API's. This was mainly done using the ChemIDplus (SRC, Inc.) database using the webchem R-package [5] and the CACTUS NCI/CADD-database using the chemtrans MATLAB function [35]. As a consequence of this, certain CAS-entries might deviate from those of for example SciFinder. However, all compounds are accurately described by the given CAS-numbers in the used databases. It should be noted here, that some compound CAS-numbers represent sodium salt anhydrides or HCl stabilised compounds.

It is the aim that most of these compounds will be used in identification-only analysis, where the primary objective is to identify each compound. It would in principle also be possible to estimate detection limits by semi-quantitative analysis by employing either surrogate analysis in representable samples or by spiking the compounds into the sample matrix. This will be pursued if time restrictions in the project allows for it.

The standards are planned to be combined into several different mixtures depending chemical properties and required solvents. This enables easier laboratory handling, concentration control, and storage.

By the end of August 2020, certain compounds are still under delivery from producers. Depending on import restrictions, the compound list is therefore still bound to change. The list is subject to change if certain compounds are incompatible as analytical standards - either due to water insolubility or instability – or if characteristic compound isomers are discovered.

Covered legislations

To ensure that the selected compound list covered all legislated compounds mentioned in the current water monitoring programmes across the EU, several Danish and European Union directives and priority lists were consulted:

- The endocrine disruptors lists I-III (https://edlists.org/)
- Priority Substances and Certain Other Pollutants according to Annex II of Directive 2008/105/EC
- Proposal for a Directive of the European Parliament and of the Council on the quality of water intended for human consumption Annex I (2018)
- MST: Bilag 3: Oversigt MFS program
- NOVANA: Screeningsundersøgelse for humane lægemidler i vandmiljøet (2015)
- Bekendtgørelse om fastlæggelse af miljømål (nr. 1625 af 2017)
- Delprogram for grundvand – opdaterede bilag 7.1-7.3, NOVANA 2017/10 (2020/05/05)
- Screeningsliste 2020, bilag 7.3b, opdateret 2-10-2020.
- Fagligt notat om resultater af massescreening 2019. Bilag 2 – Stofliste over pesticidstoffer udvalgt til massescreening i GRUMO og i GKO 2019 (pr. 06. februar 2019).
- NOVANA: Det nationale program for overvågning af vandmiljøet og naturen 2011-2015, Programbeskrivelse 2. del
- NOVANA: Det nationale program for overvågning af vandmiljøet og naturen 2017-2021. Bilag 3.3,
- Miljøfremmede stoffer og metaller i vandmiljøet. NOVANA. Tilstand og udvikling 2004-2012. (2015)
- Miljøstyrelsen: Siloxanes - Consumption, Toxicity and alternatives. (2005)

Appendix 3. GC-EI(+)-HRMS compound annotations

Appendix 3 consists of a direct exportation of GC-EI(+)-HRMS data annotated by the deconvolution node in Compound Discoverer 3.2. All data has been filtered according to a NIST library search, total score >70%, and HRF score >70%.

It can be found as an attached Excel-file.
Here is a link to Appendix 3. The Appendix can also be found through mst.dk/service/publikationer

HITLIST2

Non-targeted screening analysis (NTS) based on high-resolution mass spectrometry (HRMS) is a novel holistic approach, for obtaining information on the chemical fingerprint of an environmental sample. This project is a systematic qualitative and quantitative validation study; identifying which chemical substances on a defined suspect list can be analyzed on which analytical NTS platforms and down to which concentration level in a groundwater sample.

A suspect list of 967 xenobiotics was constructed to contain environmental pollutants of special concern (MFS) for the aquatic environment.

Analysis of standard mixtures containing the 967 chemicals showed that in combination, five NTS platforms were able to detect 70% of the studied chemical space (679 chemicals). The combination of three of the platforms; nLC-ESI(+)-HRMS, AEC-ESI(-)-HRMS and GC-EI(+)-HRMS covered 66% of the studied compounds.

Twelve 2 L grab samples of groundwater were collected from one DGU well. Six samples were collected in glass containers and six in HDPE containers. Two samples in each container type was either left unspiked, pre-spiked (before pre-concentration) or post-spiked with the 967 chemical substances: 84% of the 679 chemical compounds detectable on the platforms were recovered on a 0.1 µg/L spike-level. The limit of detections (LOD) for most substances were below 0.04 µg/L

The results show that the combination of different NTS platforms is necessary to ensure that a broad chemical range is included in future NTS methodologies. Certain types of chemicals were identified as not being resolved in this NTS methodology.

The Danish Environmental
Protection Agency
Tolderlundsvej 5
DK - 5000 Odense C

www.mst.dk