Ministry of Environment
of Denmark
Environmental
Protection Agency

HITLIST

Holistic non-targeted approach to determine pesticide and biocide residues in the aquatic environment

Pesticide Research
no. 193

March 2021

Publisher: Environmental Protection Agency

Authors:
Martin Hansen,
Emil Egede Frøkjær,
Marie Rønne Aggerbeck,
Thorsten K. O. Gravert,
Joseph D. Martin Robert B. Young,
Katrin Vorkamp
Rossana Bossi,
Department of Environmental Science, Aarhus University

Prior to its publication, this scientific report was reviewed by eight members of the monitoring group 'Water' ('Følgegruppen for Vand').

Graphics: Department of Environmental Science, Aarhus University

Photos: Department of Environmental Science, Aarhus University

ISBN: 978-87-7038-283-0

Abstract

Recent technological developments in novel analytical high-resolution mass spectrometry equipment and advanced data processing tools have made a ‘non-targeted analysis’ concept feasible to find known and unknown environmental pollutants.

A momentous drawback of the current applied targeted chemical analysis approaches used in national environmental monitoring programmes is the exclusive focus on a pre-defined list of compounds for detection. Hence, other chemical entities, potentially also present in the given sample, are not observed and their presence are unnoticed. Non-target analysis is chemical sample analysis without any prior knowledge about its chemical content. The idea is to cover as many chemicals as possible, without focusing on a predefined selection.

This research project HITLIST demonstrates that non-target analysis can elucidate a wide range of pesticides and biocides, as well as other xenobiotics and natural substances, in a broad range of water samples. The project developed suspect and non-targeted screening approaches on commercially available technologies and solutions, enabling other research laboratories, enterprises and academia to establish such non-targeted analysis methodologies. The performed research optimized two high-resolution mass spectrometry platforms hyphenated either with liquid chromatography or ion exchange chromatography. Water samples from various sources, i.e. waterworks, groundwater wells, surface water, coastal water, wastewater effluent and rainwater were analysed by non-targeted analysis. Water samples were prepared for analysis by solid-phase extraction or direct injection. A total of 45 samples were analysed with one or both high-resolution mass spectrometry platforms within the project and in general the non-targeted analysis revealed more than a thousand substances in every sample.

Rainwater samples collected on bimonthly basis from one site were analysed with one platform. The data showed seasonal traces of several pesticide and biocide residues, e.g. azoxystrobin, metazachlor and tetraconazole. These findings could indicate long-range atmospheric transport or illicit pesticide and biocide use.

Coastal water samples from one site were analysed with one platform. More than 3,000 substances were observed across these samples and it is currently possible to identify 2% of these chemical entities (e.g. tramadol, venlafaxine and prosulfocarb).

Wastewater treatment plant effluent from three locations were analysed with one platform. Thousands of chemical entities were observed across these samples and principal component analysis showed that the chemical profiles differed notably between the three sites.

Groundwater were sampled from several sites and analysed with both non-targeted analysis platforms. Thousands of substances were observed across these samples, such as 3-chlorobenzoic acid, 2-naphthalenesulfonic acid and 3-phenoxybenzoic acid.

Drinking water were sampled at two waterworks and analysed with both platforms. Across these samples nearly a thousand substances were observed and it was possible to identify 2% of these, e.g. mecoprop and dimethachlor ESA, however also many natural occurring substances such as 3-hydroxyvaleric acid and malic acid were identified.

The non-targeted analysis concept can readily be implemented in current environmental monitoring programmes and further developments are suggested, e.g. investigation of chemical space and optimization of data processing tools. It will be possible to perform retrospective non-targeted analysis by revisiting the data archives with new information and revised data processing pipelines.

Content

Acronyms

CD	Compound Discoverer
CLAM	Continuous low-level aquatic in-situ monitoring device
DPC	Desphenyl chloridazon
EPA	Environmental Protection Agency
FISh	Fragment ion search scoring algorithm
FWHM	Full width at half maximum
GC	Gas chromatography
GRUMO	Groundwater monitoring programme
HLB	Hydrophilic-lipophilic-balanced
HRMS	High-resolution mass spectrometry
IC	Ion exchange chromatography
LC	Liquid chromatography
MS	Mass spectrum
NL	Normalised level
NOVANA	Det Nationale Overvågningsprogram for Vandmiljø og Natur (The national monitoring program for aquatic environment and nature)
NTA	Non-targeted analysis
NTS	Non-targeted screening
PE	Person equivalents in wastewater treatment
SPE	Solid-phase extraction

1. Background

1.1 The Anthropocene and xenobiotics in water

Our present time period has recently been termed the *Anthropocene* as it is majorly influenced by mankind polluting Earth and causing a major loss of biodiversity[1]. With a growing population and the intensification of agriculture, that mankind relies on for e.g. health and food production, by-products of pesticides and biocides are dispersed into the environment. Many of these substances are more toxic to other organisms than the target organisms and/or mimic natural biomolecules leading to a detrimental impact on environmental and public health. The presence of pesticide and biocide residues in the environment, especially in our drinking water supplies, is of the highest public interest. With nearly 500 active ingredient pesticides and thousands of different biocide products approved for use within the European Union, there is a need to understand if residues of these bioactive substances are present in vital environmental resources, such as groundwater and other aquatic ecosystems. Current upper limits for pesticides in water are 0.10 µg/L for individual compounds, with the sum of pesticides not exceeding 0.50 µg/L, according to the EU Drinking Water Directive[2]. Environmental monitoring programmes, such as the Danish NOVANA program[a], regularly determine pesticide concentrations in the environment. These are based on highly specialised sensitive and accurate targeted analytical methods, allowing for quantification, trend analyses and checks of compliances with politically set environmental quality standards or other threshold values.

1.2 Non-targeted analysis

A momentous drawback of the targeted analytical approach, applied in NOVANA and similar monitoring programs, is the exclusive focus on a pre-defined 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.

To overcome this caveat, the purpose of this project was to develop and apply an ultra-high-resolution mass spectrometry (HRMS) non-targeted analysis (NTA) approach to identify pesticides and biocides, and derived environmental transformation products present in the aquatic environment. The term non-target analysis is used when former unknown compounds are analysed and identified in a sample, generally without a reference standard or target list[3].

Only with the HRMS technology is it possible to identify unknown substances by providing mass, molecular formula, and tentative molecular structure. NTA is fundamentally different from targeted monitoring strategies and has an enormous potential for effective evaluation of water quality regulations[4,5]. We will use this novel scientific approach to go beyond the substances listed under the European Union Water Framework Directive and commonly monitored under NOVANA, and potentially highlight other xenobiotics of emerging concern.

The development and maturation of this state-of-the-art approach for aquatic matrices can potentially lead to an inclusion of this concept in future national pollutant monitoring programs and include additional environmental matrices (e.g. sediment and biota).

[a] Nationalt overvågningsprogram for vand og natur (https://mst.dk/natur-vand/overvaagning-af-vand-og-natur/)

Aim of project and framework

The objectives of this project are;

1. Develop a non-targeted analysis methodology
2. Apply the methodology on a wide-range of aqueous environmental matrices and provide an initial list of identified substances.

To meet these objectives, the project will cover four areas: a) Field sampling by deployment of novel active samplers, b) extraction and sample preparation in the laboratory, c) instrumental analysis by HRMS, and finally d) postprocessing by targeted and non-targeted screening of the acquired data to attain lists of identified substances.

This holistic method approach is illustrated by the overview seen in Figure 1, depicting the concept of the research project.

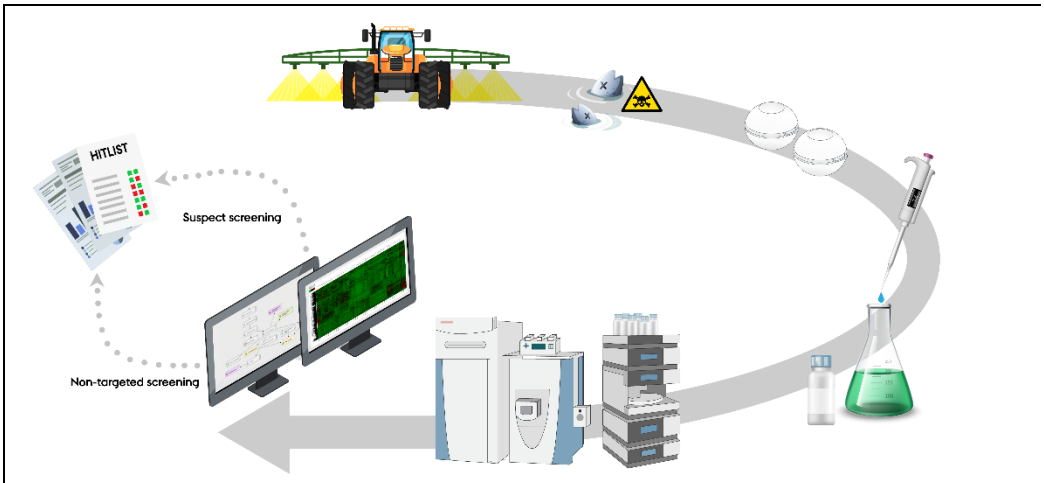

FIGURE 1. The project covers deployment of field samplers (white spheres), extraction of filters (pipette and green solvent), analysis across HRMS platforms (instrumentation) and data post-processing to identify chemical entities present in water samples.

2. Methods

Several field sampling and laboratory techniques were applied and developed in this project to optimise the NTA method to detect and identify pesticides and other xenobiotics in surface water, groundwater, drinking water and rainwater samples. The samples were collected at different sites in Denmark between 2018-2020. A number of analytical techniques and concepts were developed, optimized and applied in the present project and are described in broad terms in this chapter and more in detail in Appendix 1.

2.1 Samples and sampling sites

We performed field sampling and analysed samples from several matrices and sites (Table 1). Active samplers (CLAM) were typically extracting water at each site for several days, whereas grab samples for direct injection were brought back and processed in the laboratories.

TABLE 1. An overview of the 45 samples analysed in the project.

Type	Site	Description	Samples	Volume	Platform
Rainwater					
	A	Bimonthly collected	15	1-5 L	IC-HRMS LC-HRMS
Surface water					
	B	Coastal water site	3 CLAM†	10-20 L	LC-HRMS
	C	Wastewater effluent from 3 water treatment plants	3 grab samples in triplicate‡	0.10 L	LC-HRMS
Groundwater					
	D	Groundwater well	1 grab	0.001 L	IC-HRMS
	E	Groundwater wells (GRUMO)	8 grab‡	0.5 L	LC-HRMS IC-HRMS
Drinking water					
	F	Large waterworks (> 3M m^3/yr)	3 CLAM†, 3 grab	80-100 L	LC-HRMS IC-HRMS
	G	Small waterworks (< 0.3M m^3/yr)	3 CLAM†, 6 grab	250-300 L 0.001 L	IC-HRMS

†Continuous-low-level-aquatic-monitoring in-situ device. ‡Processed in laboratory with solid-phase extraction.

2.2 Sample preparation

Trace level concentrations of pesticides and biocides, or other xenobiotics, can lead to detection difficulties if the samples are not pre-concentrated before injection. To overcome this issue, water samples were extracted by solid-phase extraction (SPE). We used continuous-low-level-aquatic-monitoring in-situ sampling procedures[6] (CLAM) in combination with SPE (Appendix 1). This approach enabled in-situ filtering of more than 100 litres of water over several days onto a single SPE disk that was brought back to the laboratory for extraction and HRMS analysis. Grab samples from drinking water and groundwater sources were directly injected in the instrumental platforms and used as a comparison to the enriched CLAM samples. In addition, the

100 mL grab samples from wastewater effluent (sites C) and groundwater (site E) were also processed in laboratory with SPE (Appendix 1.4).

2.3 Instrumental analysis

For the highly selective identification of mass, elemental composition, and structure of unknown molecules[7], two Orbitrap-based high-resolution tandem mass spectrometry (HRMS) platforms were used in the project: A system hyphenated with nanoflow reverse-phase liquid chromatography (LC-HRMS), and a second system with ion exchange chromatography (IC-HRMS). This was used in order to cover a wider range of compounds by separation of both polar, non-polar, and ionic molecules, as well as high resolution mass spectra of both fragmented (MS/MS) and unfragmented molecules (MS). For all instrumental specifications, see Appendix 1.7 and Appendix 1.8.

2.4 Data post-processing

A two-way data post-processing strategy was used to detect compounds of environmental concern (as illustrated by Figure 2). A suspect screening was performed in the search for expected compounds of interest, and followed by an NTA pipeline used to reveal novel compounds of interests, with both methods complementing each other. The suspect screening was based on a suspect list (Appendix 2) containing 2,088 compounds compiled from different pesticide databases. The acquired data was compared automatically to mass spectrometric records for each compound in the suspect list, such as exact mass and elemental composition. The potential matches were then filtered and manually evaluated, leading to a list of compounds detected. To supplement this approach, NTA workflows were performed using the commercially available software Compound Discoverer 3 to decipher the acquired HRMS data. Using in-house pipelines, constructed from ca. 30 different nodes and hundreds of data post-processing parameters, a compound list - where each identified entity would follow the Schymanski-scale of annotation (Figure 3) - was generated for each respective measured sample. For more information regarding compound annotation, see Appendix 1.10.

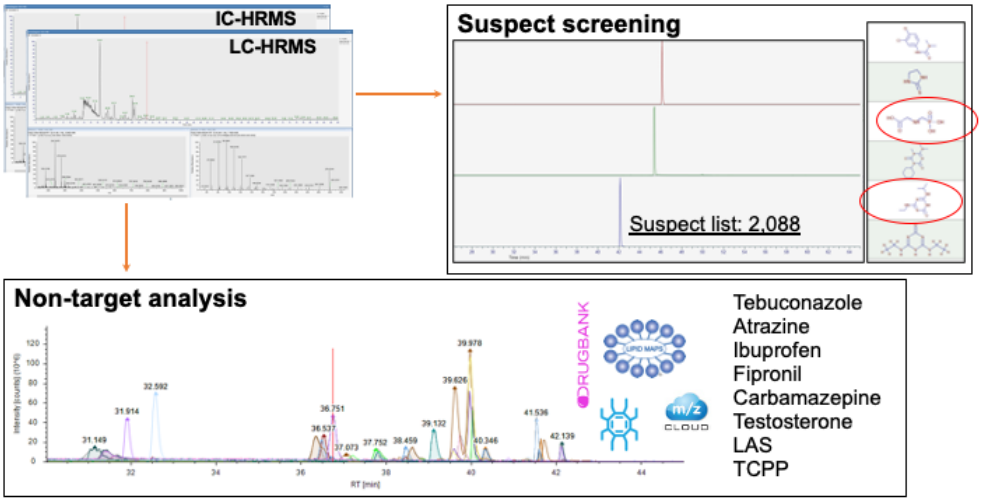

FIGURE 2. The used two-way workflow combines suspect screening and non-targeted analysis: Recorded IC and LC-HRMS data (top left) is screened for anticipated suspects (top right) and processed by NTA to discover unknown compounds (bottom).

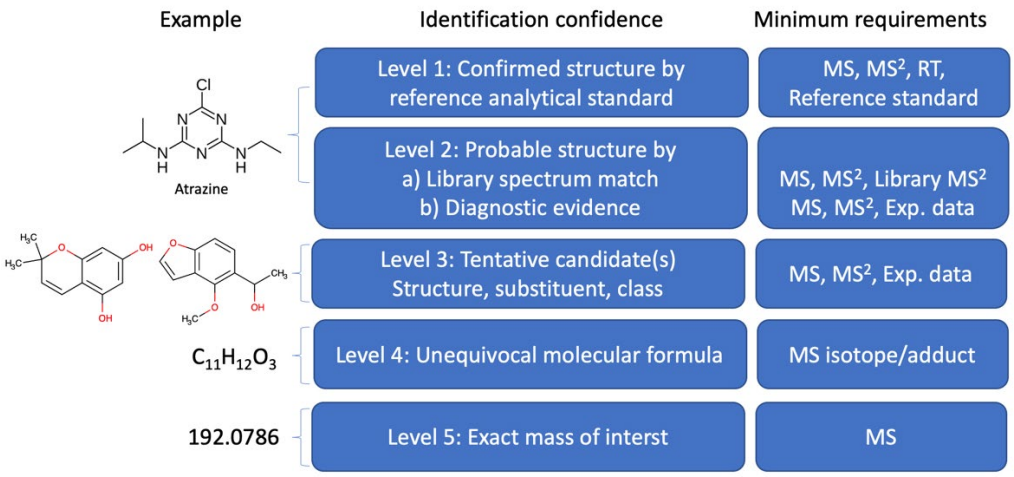

FIGURE 3. Identification confidence levels in HRMS non-targeted analysis and the minimum data requirements – the "Schymanski scale". MS and MS^2 is single and tandem mass spectrometry data, RT is chromatographic retention time data. Figure is modified from Schymanski et al.[7]

All reported annotations in this project have an identification level 1 or 2, described as follows:
Level 1: The highest level of annotation yields a confirmed structure. Compound annotation is based on full spectral and retention time match between a sample recorded MS2-spectrum and the MS2-spectrum of an in-house recorded analytical standard. In this project we recorded the spectral data of 494 pesticide analytical standards (see Appendix 6.2).
Level 2: The second-highest annotation level yields a probable structure. Compound annotation is based on a full spectral match between a recorded MS2-spectrum and a reference MS2-spectrum from a high-resolution spectral library database (e.g. mzCloud).

3. Results and discussion

3.1 Rainwater

Rainwater were obtained from an undisclosed site A in an ongoing NOVANA monitoring program. The rainwater samples were collected (up to 5 L under cooling) on a monthly basis during 2018. The samples were extracted and analysed with targeted GC-MS analysis (due to the expected volatility of suspects) to quantify 19 pesticides, PAHs and 7 nitrophenol residues in the sample extracts (Appendix 2) and the report on Atmospheric Deposition in currently under review internally at the Danish Environmental Protection Agency.

We also analysed 15 of the rainwater extracts using the NTA LC-HRMS platform. Across this dataset we discovered thousands of molecular entities and have identified (level 1 and 2) several pesticide and biocide residues, viz. azoxystrobin, metazachlor, tetraconazole, tebuconazole, flufenacet, pencycuron, prosulfocarb. Some of these compounds were also verified in the targeted GC-MS analysis, along with DNOC and 4-nitrophenol that were identified by the IC-HRMS platform. In addition, a seasonal trend is observed (Figure 4), likely due to local application of different agricultural pesticides during different periods. Some of the discovered pesticides are not approved for use in Denmark, hence they could originate from long-range transport or from illicit use (see note in Appendix 4). A list of other identified substances is found in Appendix 5.

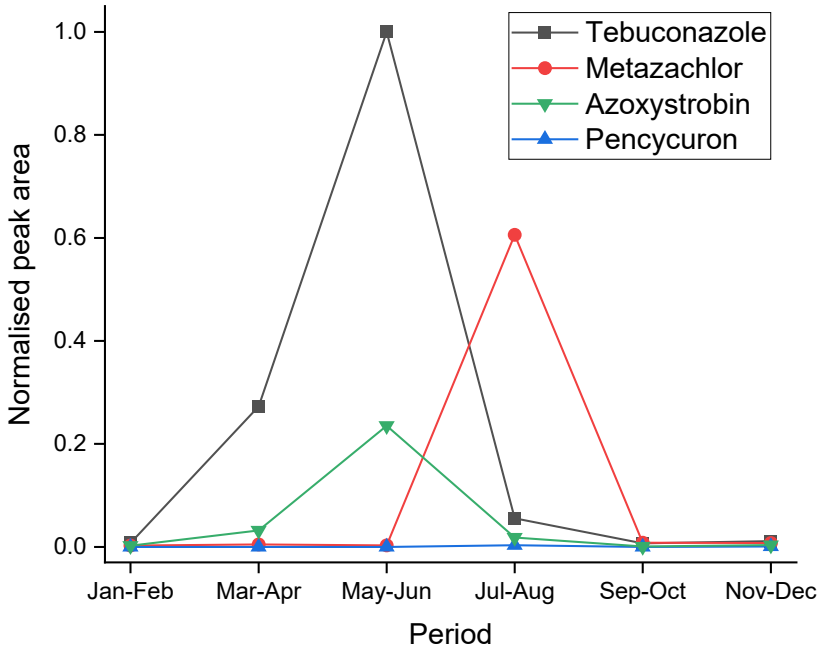

FIGURE 4. Occurrence trend of four selected substances (tebuconazole, metazachlor, azoxystrobin, and pencycuron) in rainwater collected under the NOVANA program. In March-June an increase in signal for pencycuron is seen. From spring through summer (March-June), tebuconazole and azoxystrobin is showing an increase in signal. An increase in metazachlor signal is seen during late-summer (July-August). Signals are background subtracted and normalized to highest peak intensity. September and October are averaged from weekly samples.

3.2 Surface water

3.2.1 Coastal water

Three costal water replicates were extracted with CLAM samplers at an undisclosed site B and the extracts were processed as described in Appendix 1.3 and analysed with LC-HRMS platform. More than 3,000 compounds were observed across these samples and it is currently possible to annotate 58 compounds at level 2 confidence, i.e. corresponding to a 2% identification rate. Amongst these were several xenobiotic compounds of interest: citroflex 2, tramadol, venlafaxine and prosulfocarb (Figure 5).

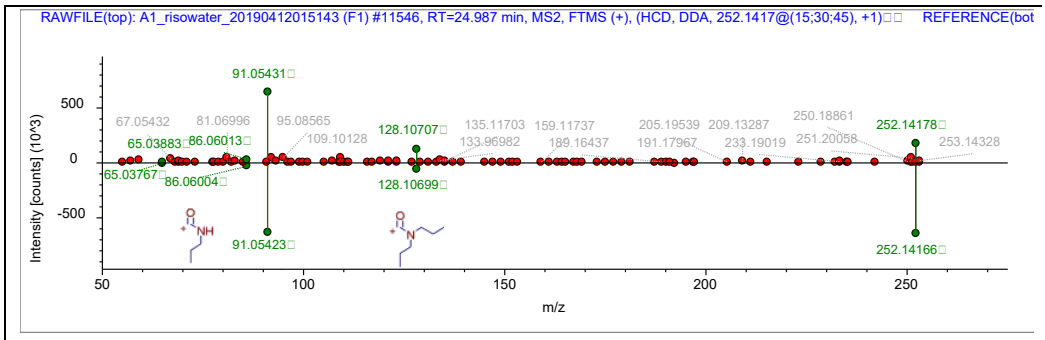

FIGURE 5. Fragmentation mass spectrum (MS2) comparison of prosulfocarb between measured spectrum in a coastal water sample (top) and the cloud-based spectral library mzCloud (bottom). Green coloured centroids confirm fragment matches. Collectively, these data resulted in a level 2 identification.

3.2.2 Wastewater effluent

We also tested the holistic NTA concept on wastewater effluent discharge from three Danish (undisclosed sites: C1, C2 and C3) water treatment plants. Site C1 is a large plant built in 1965 and retrofitted with chemical and biological treatment with a PE capacity of 345,000. It receives both industrial and residential waste. Site C2 is another large plant, receiving its wastewater from a large city, including waste from several hospitals. It has a PE of 350,000 and is built as a mechanical plant in the 1930's, with retrofitting of biological and chemical treatment. Site C3 is a small, rural treatment plant with PE 12,000, built in the 1990's, with mechanical and biological treatment processes. We collected three grab sample replicates at each site (between 8 am and 10 am during weekdays) and extracted the effluent water samples according to the procedure described in Appendix 1.4.

We observed thousands of chemical entities and of these, around 200 pharmaceuticals and transformation products were identified at level 2. Principal component analysis revealed that replicate samples were highly correlated and the chemical profiles differed notably between the three sites (Figure 6). When comparing sites, we found no immediate correlation between size, or geographical placement. Pesticides and biocides were found in every single sample; among the identified pesticides and biocides were cycloheximide, DEET, furmecyclox, propiconazole, terbutryn, verrucarol, dimefuron, pindone and prosulfocarb. The distribution of the various pesticides differed between sites, with the fewest pesticides being released in the small, rural plant (Figure 7). We also discovered a number of industrial chemicals, viz. benzotriazole, triphenyl phosphate, rhodamine 6G, benzyl butyl phthalate, centralite, dibutyl phosphate, octadecanamine, PPG-n4 and triethyl phosphate.

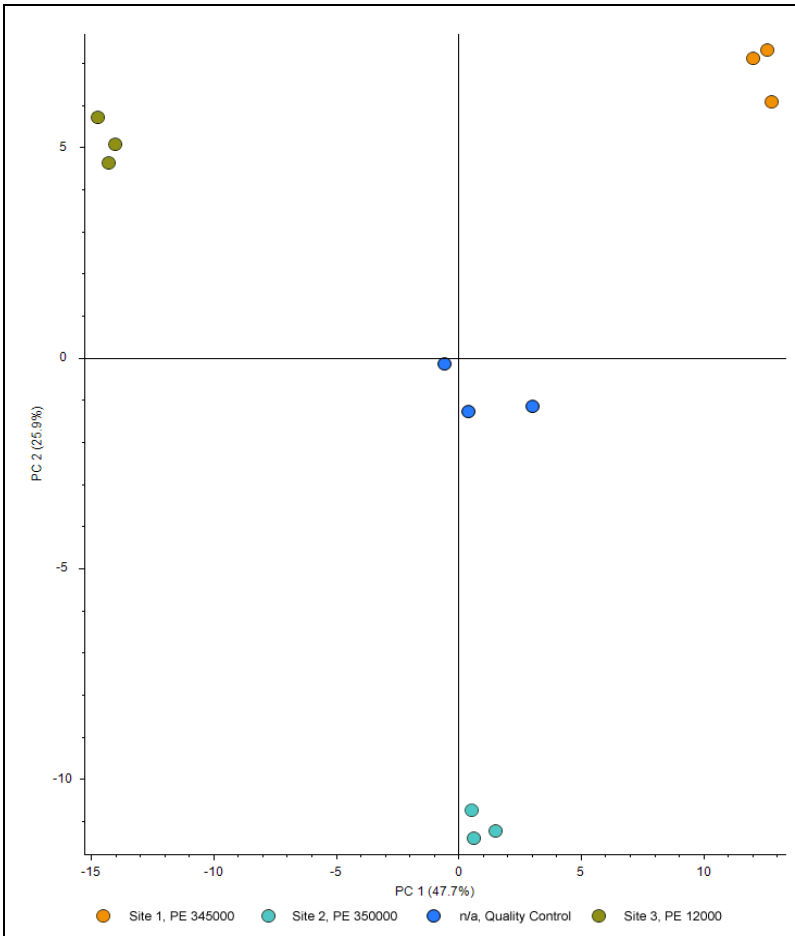

FIGURE 6. Principal component analysis of three wastewater effluents. Three replicates of each sample site. Cross-site pools were used as quality control samples (blue, as PCA centre point).

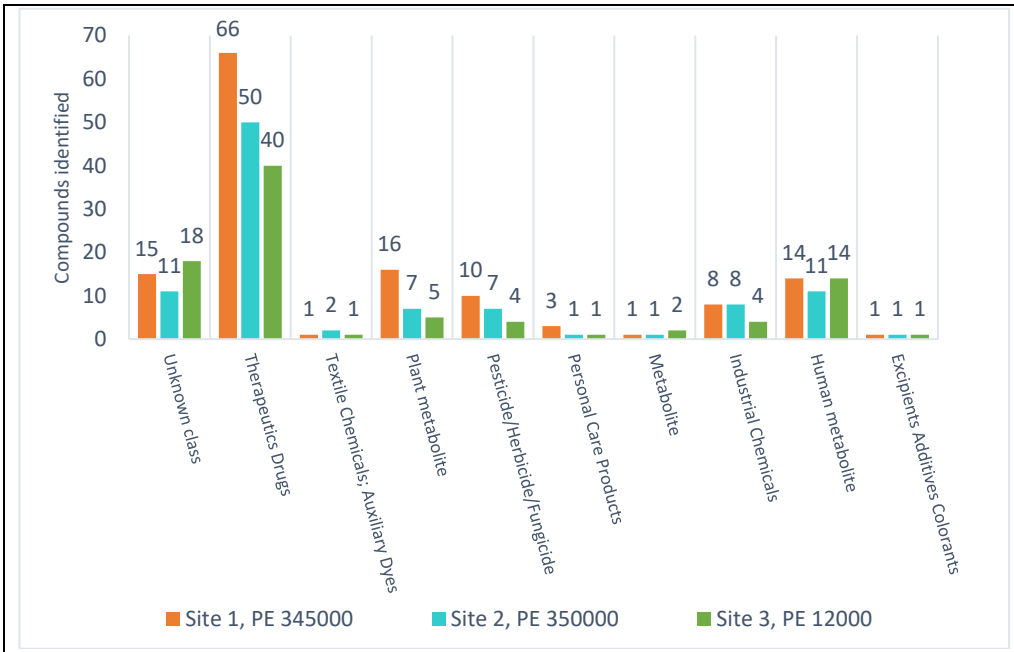

FIGURE 7. Compounds identified in wastewater effluent by NTA. Substance classes are based on Chemspider classifications (Royal Society of Chemistry).

3.3 Groundwater

Through collaboration with the Danish Environmental Protection Agency it was possible to obtain samples from groundwater wells helping to build the foundation of the post-processing workflows for analysis of water samples.

3.3.1 Groundwater well

A grab sample from a groundwater well site D for direct analysis was analysed by IC-HRMS. The initial analysis showed 1597 unique features and a deeper processing workflow narrowed these features down to 14 annotated compounds (level 2). Amongst these were the following compounds of interest: levulinic acid, 2-naphthalenesulfonic acid, 3-hydroxydecanoic acid, and 2-AEP, as all these compounds – with the exception of 2-AEP - are possible xenobiotic metabolites. 2-AEP does however have a structure very similar to that of AMPA – a known metabolite of glyphosate – with the possibility that future research could reveal a link between these two molecules.

3.3.2 Groundwater GRUMO

We analysed eight groundwater grab samples from the GRUMO program using both platforms (LC- and IC-HRMS). These samples are in this report collectively named Site E. Prior to analysis the samples were purified with solid-phase extraction as described in Appendix 1.5. An NTA workflow revealed more than 15,000 chemical entities across this dataset and from our post-processing pipelines we currently have identified 47 compounds at confidence level 2 (Appendix 6.1).

In addition to NTA, a suspect screening of five pesticide metabolites (1,2,4-triazole, desethyl desisopropyl atrazine (DEIA), desisopropyl atrazine (DIA), desphenyl chloridazon (DPC), and methyl desphenyl chloridazon (MDPC) provided by GRUMO) revealed a weakness in the HRMS data-acquisition protocol. By only fragmenting the MS-features with highest intensities, trace level compounds were omitted from the fragmentation (MS2) data disabling any further identification of these. An alternative way of recording data should be included to generate NTA data covering trace molecules. Proposed methods are the implementation of an acquisition suspect list that ensures the fragmentation of detected masses of interest following an iterative acquisition method[8].

3.4 Drinking water

The aim of these studies was to test and implement the NTA platform on drinking water samples from two different waterworks to detect pesticides and other xenobiotics in low concentrations (<0.1 µg/L), and to test the post-processing platform when using highly pre-concentrated samples (CLAM).

3.4.1 Large waterworks

Three CLAM and three direct analysis grab samples were collected from a large-scale waterworks capable of delivering >3M m^3 drinking water per year. These samples are in this report collectively named ‘Site F’. All samples were eluted following the procedure described in Appendix 1.3, analysed with IC-HRMS, and data were post-processed (Appendix 1.9). A combined amount of >3000 features were reduced to a total of 81 compounds (at level 2). The data additionally revealed two pesticides at level 1 (mecoprop and dimethachlor ESA). These two pesticides were detected in all three CLAM-sample triplicates, however neither were seen in any of the directly injected grab samples. By spiking the sample extracts with dimethachlor ESA its presence was confirmed. This is described in detail in Appendix 6.2. Figure 8 shows both mzCloud (level 2) and in-house (level 1) reference spectra of the measured MS2-spectra for dimethachlor ESA (a figure showing the same for mecoprop can be seen in Appendix 6.2). Using the preconcentration sampling method provided by the CLAM-samplers enabled accurate identification of two pesticides present in the drinking water from site F.

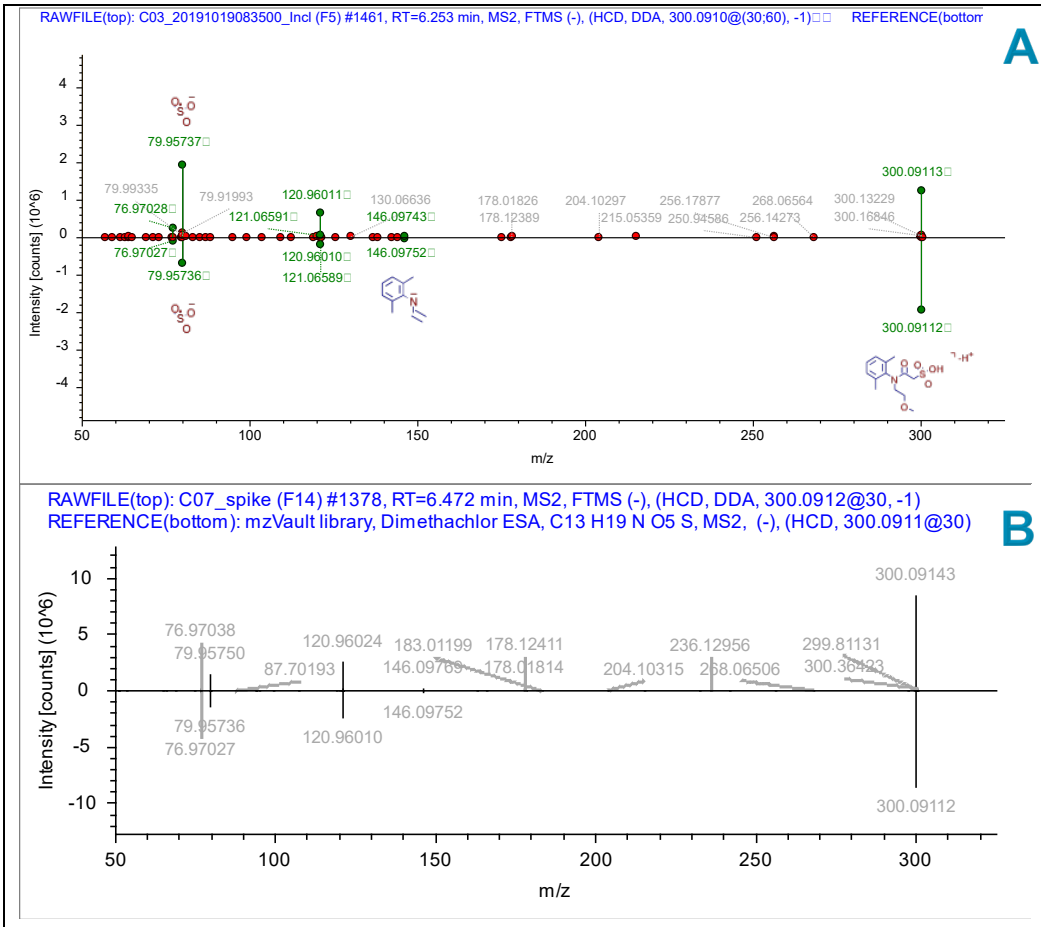

FIGURE 8. (A) Comparison of measured MS2 spectrum of dimethachlor ESA (top) (at m/z = 300.0910) in CLAM samples measured by IC-HRMS with (bottom) spectral library (mzCloud) of dimethachlor ESA with a spectral match of 95.0%. (B) Comparison of measured MS2 spectrum of dimethachlor ESA (top) (at m/z = 300.0910) in CLAM samples measured by IC-HRMS with (bottom) in-house library (mzVault) of dimethachlor ESA with a spectral match of 96.3%.

3.4.2 Small waterworks

Three CLAM and six direct analysis samples were collected from a small waterworks capable of delivering around 0.3M m^3 drinking water to nearby households per year. These samples are in this report collectively named 'Site G'. These were eluted as described in Appendix 1.3, analysed by IC-HRMS, and processed as described in Appendix 1.9. Across the dataset, we identified 2800 features where 23 compounds could be annotated at confidence level 2. The majority of these 23 substances corresponds to smaller (MW 100 - 200 Da) naturally occurring acidic metabolites such as 3-hydroxyvaleric acid, 10-HDA, and malic acid. A few could be categorised as potential xenobiotics, such as 2-naphthalenesulfonic acid and levulinic acid – naturally occurring compounds also detected in a groundwater well (site D). No evidence suggested the presence of pesticides in the samples. This evidence was supported by a spiking experiment (described in Appendix 1.6) showing that 25 out of 48 possible pesticides were detected at concentrations <0.1 µg/L. This meant the platform was able to resolve several pesticides, even though none were detected in the samples. These numbers would still suggest a risk of obtaining false negatives, an overall compound detection rate of 50% using an NTA approach validates its proof-of-concept and leads to a good baseline for this protocol. The spiking results are described in further detail in Appendix 6.2.

This data shows the promising possibility of going more than 100,000 times below the current lower limits of 0.1 µg/L using the current NTA pipeline. With improvements in the data acquisition, this number could likely become even higher. Additionally, various filter materials (ionic, C18 etc.) and analysis techniques (LC and GC) could make it possible to cover a much wider

range of compounds than what is simply detected by IC-HRMS, instigating an interest for further research and development of the NTA approach.

3.5 Confidence, false negatives, and false positives

The current pitfall by using the described NTA method and data processing workflows is that only full matches in spectral libraries (e.g. mzCloud) are annotated. It is possible that targets of interest are unable to be annotated due to either I) low data quality, II) lack of reference libraries, III) inappropriate sample preparation and analytical procedures, IV) concentrations below the limit of detection, V) inappropriate post-processing workflow, and VI) target mass not being recorded in MS2. Many of these issues can be avoided by the following workflow optimisation:

- Record data using an iterative approach.
- Record data using an inclusion list of suspects, such as a list of known pesticides as described in Appendix 2.
- Record appropriate standards alongside the sequence for representative in-house library entries.

These three steps should be included during future runs to ensure full MS2-data acquisition of potential suspected chemicals.

Additionally, there is the possibility of false positives. Throughout the post-processing of the many different samples, certain flaws were noted with the NTA work flow: Primarily focused on the risk of false negatives and false positive annotations. False negatives can be significantly reduced by increasing the instrumental detection sensitivity by extending the acquisition method towards the use of iterative acquisition, by recording high quality in-house libraries, and by investigating the chemical space surrounding the sampling and analytical techniques of suspect compounds. Regarding false positives, during the post-processing of the data from different sites, a few issues were noted in regards to the automated ability of the data processing workflows to annotate and identify compounds: I) Annotation of background compounds and II) incomplete database entries.

I) Annotation of background compounds

When processing data from IC-HRMS a commonly identified compound was acamprosate. Upon manual inspection of the chromatograms and corresponding mass spectra, this annotation seemed to originate from an artefact within the pipeline: A background feature at m/z = 180.041 (similar to the mass-to-charge ratio of acamprosate) was seen throughout the whole chromatographic run. The auto processed annotation of acamprosate was consistently determined at retention times corresponding to the dwell time of the IC column. The presence of acamprosate in the CLAM-samples was fully rejected based on a spiking experiment, Appendix 1.6, that – despite showing near-identical MS2 spectra between the ‘unknown’ acamprosate annotation and the annotation of the acamprosate standard – showed that the two acamprosate annotations had eluted more than three minutes apart, meaning that the non-targeted annotation of ‘unknown’ acamprosate was incorrect.

When processing the IC-HRMS data from Site F, the presence of mecoprop was initially discarded. It appeared that the data processing workflow identified mecoprop in the field blank at a 100 times lower intensity than that of the samples. Upon further inspection this was seen as an error, as no related mass-spectra could be obtained for mecoprop in the raw-file of the field blank – the compound annotation was based entirely upon a peak-fitting performed on the mass-spectrum noise. It is possible that the small traces of annotated mecoprop were artefacts from either 1) improper peak-alignment in the post-processing workflow or 2) an effect from carry-over in the IC-HRMS system. More research is needed to effectively prevent the annotation of background compounds.

II) Incomplete database entries

Another observed issue with the current NTA post-processing pipeline was from the online mzCloud database itself. When the database is incomplete it becomes possible for the data processing workflow to annotate a compound incorrectly if the only database reference is for a different molecule of same molecular formula and similar structure. This was seen when analysing the rainwater samples (Section 3.1) where a compound annotation of uracil ($C_4H_4N_2O_2$) was found at confidence level 2. Compared with the in-house reference spectrum of a pesticide with identical molecular formula: maleic hydrazide ($C_4H_4N_2O_2$), it was unclear which of the compounds that were present – as their retention times and MS and MS2-spectra were almost identical. Given the slight acidity of maleic hydrazide (pKa = 5.62)[b] and slight basicity of uracil (pKa = 9.45)[b], it is likely that maleic hydrazide would be better resolved in the IC-HRMS system. However, it is difficult to predict which of the two compounds – if not both – would be present in the samples without a more concise dataset. It is therefore uncertain whether it is uracil and/or maleic hydrazide present in the rainwater samples. The only way to confirm these annotations would be by measuring individual analytical standards and compare their respective retention times and fragmentation patterns – or to compare with already established mass spectra from online databases.

To summarise:
A general advice when using post-processing pipelines is that an analyte must be present in every replicate sample at similar peak intensities and retention times (within a certain threshold) and not observed in any blanks. Additionally, the pipeline settings should be implemented in such a way, that the annotated compound can be at a (minimum of) confidence level 2 with a similar MS2-spectrum to those in databases, as well as having a sufficient isotopic pattern match (>70%). Currently, it is not likely that data can be fully auto-processed or unsupervised without the need of manual verification by qualified personnel due to a small chance of false positives and false negatives.

3.6 Applicability of holistic non-targeted screening in national monitoring programs

Water from several sources, such as rainwater and drinking water, were investigated as a part of the project. Generally, across the studies for each matrix we are able to untangle thousands of chemical entities in the samples. However, at the present stage we can only annotate a few percent of these substances. Future studies should focus on quality assurance and harmonize methods used in NTS, i.e. investigate the chemical space and detection limits actually being captured by these platforms. Such experiments could entail using spike-recovery approaches with many substances (>1000) covering a broad range of chemical properties (e.g. carboxylic acids, phenolics, amino acids and halogenated substances).

[b] https://pubchem.ncbi.nlm.nih.gov/

4. Conclusions and perspectives

Water from several sources, such as rainwater, coastal water, surface water, wastewater, groundwater and drinking water, were investigated as a part of the project. Generally across the studies for each matrix, we are able to untangle thousands of chemical entities within each sample. However, at the present stage we can only annotate a few percent of these substances at level 1 and 2, and more research is needed in this area. This research project demonstrates that the holistic non-targeted screening concept is highly applicable in environmental monitoring programs.

Our main conclusions, experiences and perspectives are;

- The IC- and LC-HRMS platforms are tailored for NTA and each covers a different range of chemicals. The results of this study have shown the importance of applying complementary analytical platforms to cover the different chemicals, though the LC-HRMS seemed to cover a much larger range than the IC-HRMS platform.
- NTA in combination with powerful data post-processing pipelines could be developed into a routine environmental pollutant monitoring tool.
- HRMS data acquisition requires an extensive iterative protocol to ensure fragmentation of all known and unknown targets of interest for highest levels of identification confidence. Thus, a suspect list should be included in future analyses to ensure a full MS2-data acquisition of potential suspected trace level chemicals.
- More research is needed especially within cheminformatics and improvement of data post-processing methods.
- Stacking extraction disks with various solid-phase materials will broaden the extraction of the chemical space from water matrices.
- It is anticipated that the financial costs for developing new or using existing targeted analytical methods, for the xenobiotics identified in this project, is most likely higher than using NTA.
- It is our experience that new molecular information is uncovered when digging deeper and deeper into the already recorded NTA-data. Hence, retrospective analysis, revisiting the data archives with new information and revised pipelines can allow for the elucidation of new chemicals.
- The data presented in this report is a ‘first stab’. We expect to further develop data acquisition and post-processing pipelines, in combination with machine learning methods, to discover substances that are currently located in deep within the data.
- It would take an immense amount of resources to single-handedly identify the majority of substances in a water sample (thousands of chemical entities). Typically, HRMS spectral libraries in combination with chromatographic retention information are sufficient to confirm a substance. However, a single laboratory typically only builds spectral libraries with a few thousand chemicals. Consequently, there is a need to collaborate on a global scale and share HRMS spectral information. Through scientific collaboration such as the NORMAN network Digital Sample Freezing Platform and the NORMAN network Suspect List Exchange[9], it would be more feasible to elucidate the majority of chemical entities in a given sample.
- Another intriguing approach would be to use effect-directed analysis in combination with NTA. By screening the sample extract, or fractions of the sample, against a panel of in vitro toxicology assays (e.g. for estrogen receptor activity) we would be able to

highlight bioactive regions of the sample and afterward concentrate the NTA resources within this chemical space[10].

- Evidence from a spiking experiment suggested that the NTA workflow is capable of identifying compounds present in water in concentrations of 0.001 – 1 ng/L, as long as these compounds are retained in the solid-phase extraction filter.
- NTA is fundamentally different from targeted monitoring strategies and has an enormous potential for a more effective evaluation of water quality regulations.

5. References

1. Hayes, T. B. & Hansen, M. From silent spring to silent night: Agrochemicals and the anthropocene. *Elem Sci Anth* **5**, (2017).
2. Dolan, T., Howsam, P., Parsons, D. J. & Whelan, M. J. Is the EU drinking water directive standard for pesticides in drinking water consistent with the precautionary principle? *Environ. Sci. Technol.* **47**, 4999–5006 (2013).
3. Escher, B. I., Stapleton, H. M. & Schymanski, E. L. Tracking complex mixtures of chemicals in our changing environment. *Science (80-.).* **367**, 388–392 (2020).
4. Hollender, J. *et al.* High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environ. Sci. Eur.* **31**, (2019).
5. Samanipour, S. *et al.* Machine learning combined with non-targeted LC-HRMS analysis for a risk warning system of chemical hazards in drinking water: A proof of concept. *Talanta* **195**, 426–432 (2019).
6. Coes, A. L., Paretti, N. V., Foreman, W. T., Iverson, J. L. & Alvarez, D. A. Sampling trace organic compounds in water: A comparison of a continuous active sampler to continuous passive and discrete sampling methods. *Sci. Total Environ.* **473**–**474**, 731–741 (2014).
7. Schymanski, E. L. *et al.* Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology* **48**, 2097–2098 (2014).
8. Hoopmann, M. R., Merrihew, G. E., von Haller, P. D. & MacCoss, M. J. Post Analysis Data Acquisition for the Iterative MS/MS Sampling of Proteomics Mixtures. *J. Proteome Res.* **8**, 1870–1875 (2009).
9. Alygizakis, N. A. *et al.* NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in “digitally frozen” environmental samples. *TrAC - Trends Anal. Chem.* **115**, 129–137 (2019).
10. Rosenmai, A. K. *et al.* Effect-based assessment of recipient waters impacted by on-site, small scale, and large scale waste water treatment facilities – combining passive sampling with in vitro bioassays and chemical analysis. *Sci. Rep.* **8**, 1–11 (2018).
11. Ensminger, M. P. *et al.* Continuous low-level aquatic monitoring (CLAM) samplers for pesticide contaminant screening in urban runoff: Analytical approach and a field test case. *Chemosphere* **184**, 1028–1035 (2017).
12. Liu, R. *et al.* Pharmacokinetics, bioavailability, excretion, and metabolic analysis of Schisanlactone E, a bioactive ingredient from Kadsura heteroclita (Roxb) Craib, in rats by UHPLC–MS/MS and UHPLC-Q-Orbitrap HRMS. *J. Pharm. Biomed. Anal.* **177**, 112875 (2020).

Appendix 1. Methodologies

Appendix 1.1 Continuous Low-level Aquatic Monitoring (CLAM) sampling

Trace level concentrations of pesticides and biocides, or other xenobiotics, can lead to detection difficulties if the samples are not pre-concentrated before injection. To overcome this issue, solid-phase extraction (SPE) can be performed on the water samples to improve the detection limits of the analysis.

To decrease the loss of analyte during SPE, active samplers (CLAM) were used to obtain highly enriched water samples (Figure 9) in the field. Conceptually these active samplers are able to extract up to hundreds of litres of water spanning several days deployment directly through a solid-phase filter disk. SPE-disks are preconditioned with an organic solvent (methanol) before being deployed. It has been shown that deployment of CLAM-samplers will achieve the retention of a broader range of trace organic compounds (e.g. caffeine, metolachlor, triclosan and chlorpyrifos), by using hydrophilic-lipophilic balanced SPE material, with lower reporting limits compared to direct/grab-sampling[6,11] making this technique ideal for non-target analysis. Recoveries of >30 various trace organic compounds using the CLAM-samplers with HLB-filters are reported by Coes et al. to be around 7-127 %[6], however, this could be acceptable for identification purposes. These samplers were therefore deployed and tested as a part of the current project.

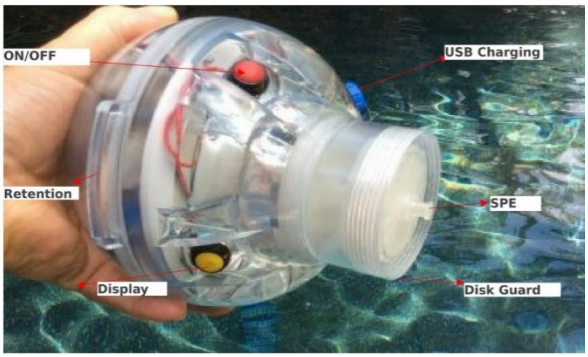

Figure 9. CLAM active water sampling unit. SPE, solid-phase extraction.

Picture reprinted from https://aqualytical.com/clam-excluding-volume-totalizer/.

Field sampling was done by deploying three parallel CLAM samplers with preconditioned 25 cm^3 hydrophilic-lipophilic-balanced polymer (HLB) SPE filter disks at each sampling site. Preconditioning of the filters was done using methanol 24 hours prior to field deployment. The filters were connected to the CLAM-sampler via clean silicone tubing (0.5 – 1 m). The CLAM-samplers could run for a duration of 2-6 days and it was possible to extract up to 300 L of continuous water sample on a single SPE-filter. After extraction, the SPE-filters were transported back to the laboratory, dried, and stored at -20 °C awaiting further sample preparation. CLAM-samplers were deployed at several sites (B, F and G). At every sampling occasion, field blanks were obtained by using non-exposed SPE-filters. Moreover, three CLAM-samplers were initially deployed in the laboratory to 1) test their performance over time, and 2) to generate equipment blanks for evaluation of the analytical background originating from the samplers. A reservoir was filled with deionised water wherein the samplers ran for a total of 48 hours.

Appendix 1.2 Site specific CLAM sampling

Site B – Coastal surface water. Three CLAM-samplers were deployed for 24 hours at an undisclosed coastal surface water site to serve as proof of concept for *in situ* active sampling by ob-

serving their field performance. A field blank (preconditioned SPE filter) was deployed simultaneously. Complications partly due to particle retainment in the filters resulted in reduced flow rate which lead to low extracted volumes (<20 L) despite running for 48 hours. Figure 10shows the formation of particles that occurred during the deployment of the SPE disks (A-C) with the field blank (D) being unaffected.

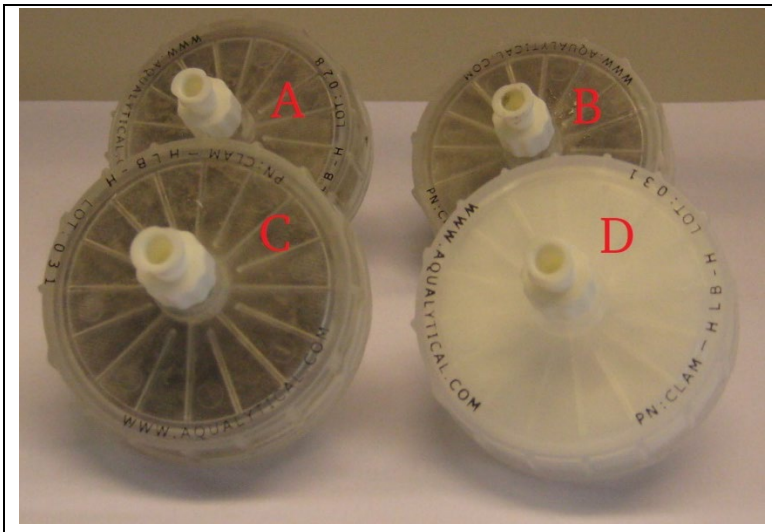

Figure 10. Four HLB SPE filter disks after being deployed for 24 hours in coastal seawater. Disks A-C are sample disks after deployment, and D is the field blank.

The three sample extraction disks were eluted according to Appendix 1.3. The primary goal for this study was to optimize the elution process, analysis platform and the sampling procedure, and to improve the post-processing pipeline for real samples.

Site F – Drinking water (waterworks). Three CLAM-samplers were deployed for 144 hours (6 days) at a state-of-the-art undisclosed waterworks facility to collect drinking water. A preconditioned SPE filter was brought to the site to serve as a field blank. Daily grab samples were taken in LC and IC vials every day at 9 am. A vial of MilliQ-water was brought along to the site to serve as a field blank for the grab samples. A steady CLAM-flow rate of ~30 mL/min (~45 L/day) was observed daily.

Site G – Drinking water (waterworks). Three CLAM-samplers were deployed for 48 hours at an undisclosed waterworks site to collect drinking water. A preconditioned HLB SPE filter disk was used as field blank. Three grab samples were collected during the first hour of the sampling. A vial of MilliQ-water was brought along to the site to serve as a field blank for the grab samples.

Appendix 1.3 CLAM sample preparation

When eluting the CLAM SPE-disks, a method intended for gas chromatography (GC) analysis was initially proposed by the CLAM-manufacturers. Several adjustments were made (based on preliminary test-runs of the CLAM-disk extraction in both laboratory and field settings) in order to make the elution procedure fit for LC and IC. The contents of the dried SPE-filters were eluted in the laboratory according to the following in-house developed procedure (Figure 11).

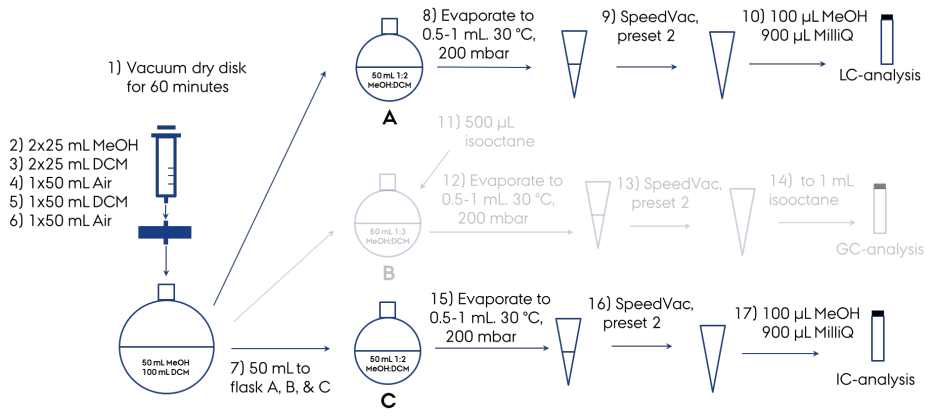

Figure 11. SPE-filter elution procedure and extract enrichment for application in LC and IC-HRMS. Methanol and dichloromethane are first injected through the SPE-filters, split into two different containers, and rotary evaporated at 30 °C 600 mbar to 0.5 – 2 mL total volume each. The remaining fractions are then slowly evaporated in SpeedVac at 65 °C and 5.1 Torr over two hours. Finally, the dried extracts are reconstituted in a solution of 10% methanol in MilliQ-water (18.2 MΩ) ready for injection in either liquid chromatography (LC) or ion exchange (IC) high-resolution mass spectrometry (HRMS)

Appendix 1.4 Wastewater effluent sample preparation

Three grab sample replicates at 100 mL were collected each site (between 8 am and 10 am) and brought back to the laboratories. The samples were extracted using solid-phase extraction (200 mg HLB) and eluted with 5 mL methanol and reconstituted in 10% acetonitrile and 0.05 % trifluoroacetic acid.

Appendix 1.5 GRUMO sample preparation

GRUMO samples were delivered cooled at the laboratories and sample volumes were approximately 500 mL. The samples were extracted using solid-phase extraction (200 mg HLB) and eluted with 5 mL methanol and reconstituted in 10% methanol, 2% acetonitrile, and 0.05% trifluoroacetic acid.

Appendix 1.6 Extract spiking

To enhance the confidence of compound detection in IC-HRMS, aliquots of sample extracts from Sites F and G were spiked with various pesticides to enable comparison of retention times and extracted mass of the identified compounds.

Extracts from site F were spiked to 5 mg/L dimethachlor ESA and 0.5 mg/L acamprosate: To 200 µL of each extract (CLAM triplicates, grab sample triplicates, and corresponding field blanks) was added 10 µL of a 100 mg/L dimethachlor ESA standard solution and 10 µL of a 10 mg/L acamprosate standard solution to approximate spiked concentrations of 5 mg/L and 0.5 mg/L respectively. The high concentrations would ensure strong intensities for better compound identification in the NTA workflow.

Extracts from site G (CLAM triplicates, grab samples from day 1-3, and corresponding field blanks) were spiked to 0, 0.1, 1, 10, and 100 µg/L respectively with a mixture of 494 pesticide standards (see Appendix 6.2) for analysis in IC-HRMS: 70 µL sample extract was added to a well plate and added 20 µL of a 0.5 (or 0.05, 0.005 etc.) mg/L pesticide standard solution (494) and 10 µL 10% methanol.

Appendix 1.7 Liquid chromatography high-resolution tandem mass spectrometry (LC-HRMS)

For LC-HRMS, a high-field Orbitrap tandem mass spectrometer (Q Exactive HF, Thermo Scientific) was used, enabling an ultra-high mass resolving power (400,000 at m/z 100) capable of performing controlled high-energy collision fragmentation. This was used to obtain high resolution mass spectra of both a fragmented (MS/MS) and unfragmented molecule (MS). The mass spectrometer was hyphenated with a LC system (nanoflow ultra-high-pressure liquid chromatography, Ultimate 3000, Thermo Scientific). Data was recorded in an untargeted analysis approach in the data dependent acquisition mode (ddms2). The autosampler and columns were thermostated at 8°C and 40°C, respectively. An amount of 20 µL sample was loaded on a preconcentration trap (C18, 300 µm x 5 mm, 5 µm, 100 Å cartridge) and eluted onto an analytical column (75 µm, 2 µm C18, at two lengths 250 or 750 mm) with a chromatographic triple-phasic gradient ranging from 10 to 95% mobile phase B (98% acetonitrile and 0.1% formic acid) at a 300 nL per minute flow rate. The HRMS system was equipped with an EASYspray ion source operated at a spray voltage of 1.50 to 2.50 kV, a capillary temperature of 250°C, S–lens RF level at 50 V and probe heater temperature at 350°C. The instrument was operated at a scan range of m/z 75−975 for full scan at a mass resolution of 240,000 FWHM (at m/z 200) and automatic gain control target of 1,000,000 ions. The maximum ion injection time was set to 50 ms. Data-dependent acquisition mode was set to trigger the top 20 most intense parent ions for MS/MS experiments and record the ion fragments at HRMS as well with a dynamic exclusion time of at least 2 s. External mass calibrations were carried out weekly, using a certified mixture of caffeine and Ultramark 1621.

Appendix 1.8 Ion exchange chromatography (IC-HRMS)

We used a reagent-free anion exchange chromatography system (ICS-6000, Thermo Scientific) hyphenated with a high-field Orbitrap tandem mass spectrometer (Q Exactive HF, Thermo Scientific). The autosampler and columns were thermostated at 8 °C and 40°C, respectively. We injected 10 µL water sample, or extracts of water, at 0.45 mL/min on to an IonPac analytical column (2 x 250 mm, 4 µm, AS19, Thermo Scientific). The analytes were passed through a conductivity detector and mixed with isopropanol via a tee-piece before infusion into the mass spectrometer. This system can resolve a large number of inorganic anions and organic acids. MS-settings were similar to those of LC-HRMS primarily focusing on a negative ionisation source. The mass spec was equipped with an HESI-II ion source (Thermo Scientific) operated at a spray voltage of 3.50 kV, a capillary temperature at 250 °C, S–lens RF level at 50 V and probe heater temperature at 350 °C. The Orbitrap was operated at a mass resolution of 240,000 at m/z 200 with a target of 3e6 ions and a maximum injection time at 100 ms, and the 5 most intense ions were selected for MS/MS fragmentation in subsequent scans. The selected ions were isolated at a m/z 0.4 window and higher-energy collision dissociation was done at 30 NCE and fragments recorded in centroid mode at a resolution of 30,000 with a 100 ms max filling time and target of 1e5 ions.

Appendix 1.9 Data post-processing

We used the commercially available software Compound Discoverer 3 to decipher the acquired HRMS data. By generating in-house pipelines, constructed from ca. 30 different nodes and few hundred data post-processing parameters, a list of ion features (one chemical substance can form several ion features, such as sodium and potassium adducts during the MS-ionization process) is derived and converted into a compound list. Typically, we generated a new pipeline for each new sample batch and have in total constructed ca. 20 different pipelines in this project. A large part of the post-processing was assigned to testing out and modifying existing work flows to improve the compound identification – for which the ideal work flow is yet to be confirmed. Basic settings, node descriptions, and work flow modification tools can be found in the Compound Discoverer v3.1 User Guide.

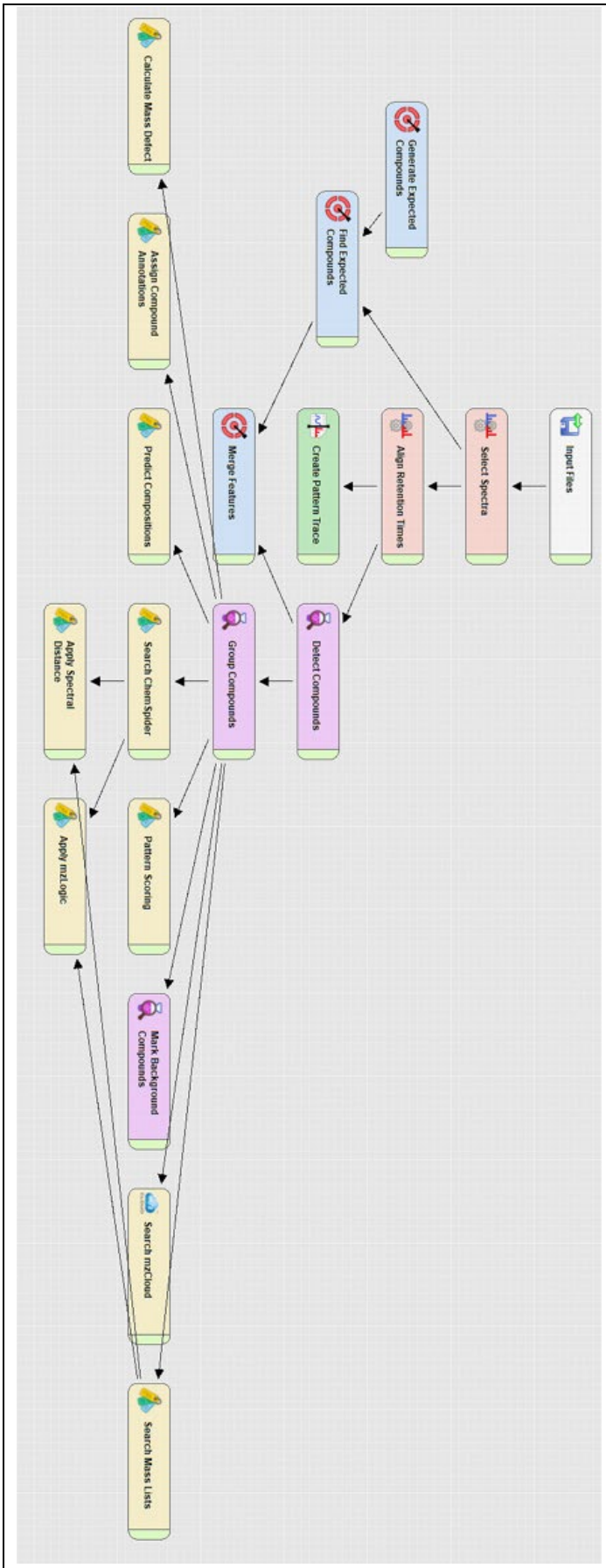

Figure 12. In-housing constructed data post-processing pipeline. Each box is a node for processing data such as "Detect Compounds" from the acquired HRMS-data files.

Appendix 1.10 Compound annotation

To achieve a compound annotation fulfilling the requirements for a level of confidence of 2 on the Schymanski scale[7] (Figure 3) every identified compound must have a recorded MS2-spectrum identical to a reference spectrum from a certified database. Using the Compound Discoverer 3 software, measured HRMS spectra is compared with the mzCloud database, which currently hosts >7 million spectra for more than 18,495 compounds. To supplement this approach, structure elucidation was further improved from *in silico* FISh scores (Fragment Ion Searching). FISh-scores are theoretically calculated fragmentation patterns of a defined structure of a presumed compound[12]. A comparison is then made with the measured MS2 spectrum yielding a value between 0 – 100, where 100 represents a complete fit between the theoretically expected fragmentation pattern of the presumed compound and that observed for the actual unknown compound. The primary identification procedure is using the mzCloud database as a reference to measured MS2 spectra. An example of this is shown in Figure 13, where the measured MS2-pattern of the molecule *N*-acetylneuraminic acid (top) is compared with the reference spectrum from mzCloud (bottom). When only a single compound can be matched from the database spectra, mzCloud defines this as a full match. When an unknown compound is identified with a full match in mzCloud, it is annotated as this compound. To improve the accuracy of the annotation a FISh-score is applied and a predicted composition is calculated. The predicted composition aims to calculate an appropriate chemical formula based upon the measured compound mass. A mass error (Δmass) is calculated by:

$$\Delta\text{mass} = \frac{(m_{\text{obs}} - m_{\text{theo}})}{m_{\text{theo}}} \cdot 10^6$$

If it exceeds 2 ppm the annotation will be rejected. A fit is made by comparison of the isotopic patterns estimated from the predicted composition, and the experimental isotopic pattern. This fit is given a value between 0 – 100, where 100 represents a complete match. By screening each annotated compound from its predicted composition, Δmass, isotopic pattern fit, spectral match, and FISh-score, it is either accepted or rejected. The aim is typically to achieve a predicted composition and mass error <1 ppm, and having a pattern fit, spectral mass, and FISh-score as close to 100 as possible. Due to the possibility of background signals and other spectral noise (etc. differences in instrumental settings between the recorded and reference spectra), the general threshold is considered to be >50 in order to account for algorithmic accuracy between the 'clean' MS2 library data obtained from pure standards, and MS2 data obtained from sample mixtures with larger matrix effects. Ensuring the appropriate threshold values is a manual process that requires sufficient pre-filtering of the dataset. Typically, this was done by disregarding any compound with an annotation confidence of less than 2. Finally, the annotated compounds are compared with the suspect list of >2000 potential xenobiotics and pesticides, fungicides, and herbicides, and their corresponding metabolites. Level 1 annotation was only possible for compounds where an in-house spectral reference was available in addition to a mzCloud match.

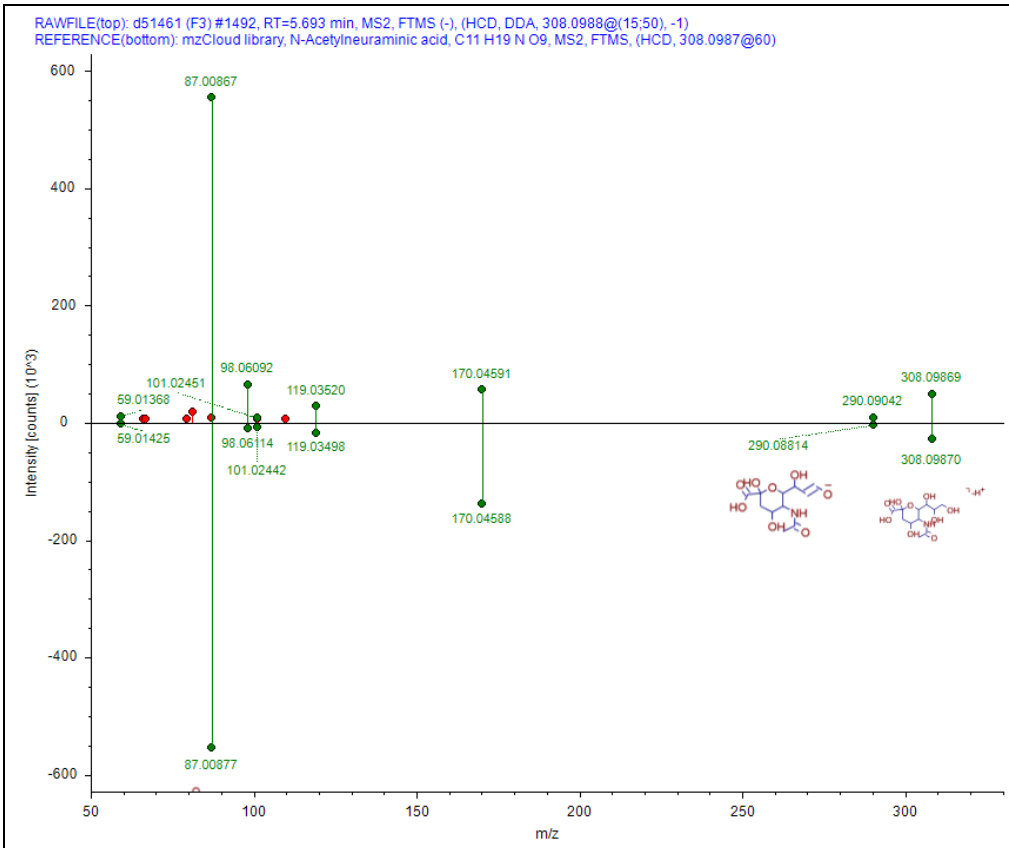

Figure 13. Comparison of experimentally obtained (top) MS2-spectrum of *N*-Acetylneuraminic acid in LC-HRMS with the recorded reference (bottom) spectrum from the mzCloud-database using Compound Discoverer version 3.1.0.279. Spectral Match score is 93.4%.

In-house HRMS spectral library and annotation

To ensure highly reliable compound identification, it was necessary to compare the sample mass spectra to a spectral in-house library in addition to the already established mzCloud library. We prepared a high-grade analytical standard library containing 494 pesticide standards at 1 mg/L in 40 % acetonitrile (see Appendix 4) analysed with LC-HRMS and IC-HRMS. The spectrum library was constructed using the software mzVault (v2.3) by extracting one MS2 spectrum for each detectable precursor in the pesticide standard solution within a mass tolerance of 2 ppm and a minimum normalised level of 100,000 (approximately a factor 10^4 lower than the top intensity). The spectra were corrected for possible measurement errors (applying the so-called recalibration algorithm with a mass tolerance of 5 ppm) and were corrected for noise. The obtained matches were then manually reviewed to decrease the number of false positives comparing the results with those of measured instrumental blanks. 69% (339) of the 494 pesticide standards were identified on the two platforms; 48 compounds were resolved in the IC-HRMS procedure, and 291 in the LC-HRMS procedure with 7 compounds being identified by both methods. This implies that though LC-HRMS can resolve a wide range of compounds, IC-HRMS is a useful – and unique - complementary tool to further increase this range of detected compounds. Additionally, this could be evidence that there is still need for another complimentary method (GC, LC(-), or IC(+)-HRMS) and workflow optimisation in order to resolve the remaining library entries. This should be discovered in a future project. These resolved compounds were used as in-house library entities serving as reference spectra to compounds identified in the samples by each analytical technique respectively enabling an annotation confidence level of 1 according to the Schymanski scale.

Appendix 2. Mass list for suspect screening and in-house spectral libraries

The following chemical lists are available at www.mst.dk//media/205108/hitlist-suspectlist.pdf

Appendix 2.1 Mass list for suspect screening

A suspect screening list with 2,088 entries directly related to pesticides and other similar xenobiotics. Actual compounds of monitoring interest and some of these metabolites are contained in this list, as well as compounds that in some cases can be associated with pesticides as for example preservatives and other additives present in commercial pesticide products. All these compound entries constitute an interest in non-targeted analysis, as patterns of otherwise harmless molecules could indicate the trace presence of harmful xenobiotics. This list is continuously being expanded once new suspects are discovered.

Appendix 2.2 In-house spectral libraries

To obtain level 1 compound annotation (Figure 3), it was necessary to compare the sample mass spectra to an in-house spectral library. We prepared a high-grade analytical standard library containing 494 pesticide standards at 1 mg/L in 40 % acetonitrile analysed with LC-HRMS and IC-HRMS. The spectrum library was constructed using the software mzVault (v2.3) by extracting one MS2 spectrum for each detectable precursor in the pesticide standard solution within a mass tolerance of 2 ppm and a minimum normalised level of 100,000 (approximately a factor 10^4 lower than the top intensity). Obtained matches were then manually reviewed to decrease the number of false positives comparing the results with those of measured instrumental blanks. Of the 494 pesticide standards, 48 compounds were resolved in the IC-HRMS procedure, and 291 in the LC-HRMS procedure with 7 compounds being identified by both methods. This implies that though LC-HRMS can resolve a wide range of compounds, IC-HRMS is a useful – and unique - complementary tool to further increase this range of detected compounds.

Appendix 3. Target list for rainwater analysis 2018

Pesticides	Nitrophenols	PAH's
Atrazine	4-Nitrophenol	Acenaphthen
Clomazone	2,4-Dinitrophenol	Acenaphthylen
Desethylatrazine	2,6-Dinitrophenol	Anthracen
Desethylterbutylazine	2,6-Dimethyl- 4-nitrophenol	Benz(a)anthracen
Desisopropylatrazine	3-Methyl-4-nitrophenol	Benz(a)pyren
Dichlorprop	DNOC	Benz(e)pyren
Diuron	Dinoseb	Benz(ghi)perylen
Ethofumesate		Benz(b+j+k)fluoranthener
Epoxiconazole		Chrysen+triphenylen
Hydroxyatrazine		Dibenz[a,h]anthracen
Hydroxysimazine		Dibenzothiophene
Isoproturon		3,6-Dimethylphenanthrene
MCPA		Fluoranthen
Mechlorprop		Fluoren
Metamitron		Indeno(1,2,3-cd)pyren
Metazachlor		1-Methylnaphthalen
Pendimethalin		2-Methylnaphthalen
Prosulfocarb		2-Methylphenanthren
Terbutylazine		Naphthalen
		Perylen
		Phenanthren
		Pyren

Appendix 4. Note on pesticides and biocides in rain water

Norfluorazon is not registered in Denmark. Used as pre-emergency herbicide; possible long-range transport.

Azoxystrobin is a fungicide (20.258 kg active compound sold in 2017); it is applied to winter cereals in autumn, thus its presence in the samples is probably due to local sources.

Metazachlor is an herbicide not registered in Denmark; possible long-range transport.

Tebuconazole is a fungicide registered in Denmark (81.011 kg sold in 2017); it is applied to winter cereals in autumn, thus its presence in the samples is probably due to local sources.

Tetraconazole is a fungicide not registered in Denmark; possible long-range transport.

Flufenacet is an herbicide not registered in Denmark; possible long-range transport.

Pyraclosyrobin is a fungicide registered in Denmark (39.520 kg sold in 2017); it is applied to winter cereals in autumn, thus its presence in the samples is probably due to local sources.

Pencycuron is a fungicide not registered in Denmark; possible long-range transport.

Prosulfocarb is an herbicide registered in Denmark and mainly used in autumn (265.824 kg sold in 2017); presence in rainwater due mainly to local sources

The fungicides registered in Denmark are also applied to potatoes, vegetables and spring cereals, thus they should also be found in rainwater samples from May to August. This should confirm the local sources of these pesticides vs. long-range transport.

Appendix 5. Rainwater substance list

The 20 selected substances were discovered in rainwater from site A. Substances are annotated at level 1 or 2 using in house HRMS spectral library or public available mzCloud, respectively. Delta mass is the deviation, in ppm, of the measured mass from the theoretical mass.

	Name	Formula	MW	Δmass (ppm)	Library match (%)	Annotation level
1	Azoxystrobin	$C_{22}H_{17}N_3O5$	403.11738	1.38	88.6	1
2	Flufenacet	$C_{14}H_{13}F_4N_3O_2S$	363.06694	1.33	81.2	1
3	Clomazone[a]	$C_{12}H_{14}ClNO_2$	239.07141	0.45	97.1	1
4	Cyprodinil	$C_{14}H_{15}N_3$	225.12699	1.72	85.5	2
5	Prometryn	$C_{10}H_{19}N_5S$	241.13632	0.86	93.8	1
6	Fenpropimorph	$C_{20}H_{33}NO$	303.25633	0.37	96.9	1
7	Pencycuron	$C_{19}H_{21}ClN_2O$	328.1346	1.10	84.9	1
8	Epoxiconazole[a]	$C_{17}H_{13}ClFN_3O$	329.07347	1.06	96.0	1
9	Metolachlor	$C_{15}H_{22}ClNO_2$	283.13417	0.93	93.3	1
10	[1,1'-biphenyl]-2,2'-dicarboxylic acid	$C_{14}H_{10}O_4$	242.05806	0.62	76.5	2
11	Prosulfocarb[a]	$C_{14}H_{21}NOS$	251.13474	1.39	95.4	1
12	Tebuconazole	$C_{16}H_{22}ClN_3O$	307.14554	1.29	96.7	1
13	Metazachlor[a]	$C_{14}H_{16}ClN_3O$	277.09848	1.04	96.1	1
14	Fenpropidin	$C_{19}H_{31}N$	273.24596	1.15	94.9	1
15	Tetraconazole	$C_{13}H_{11}Cl_2F_4N_3O$	371.02184	0.83	89.4	1
16	Spiroxamine	$C_{18}H_{35}NO_2$	297.2669	0.40	95.8	2
17	Terbuthylazine[a]	$C_9H_{16}ClN_5$	229.10986	1.92	94.0	1
18	7-(2-hydroxypropan-2-yl)-1,4a-dimethyl-decahydronaphthalen-1-ol	$C_{15}H_{28}O_2$	240.20935	1.74	76.6	2
19	Prothioconazole-desthio	$C_{14}H_{15}Cl_2N_3O$	311.05966	1.43	89.7	2
20	Propyzamide	$C_{12}H_{11}Cl_2NO$	255.02224	1.86	93.1	2
21	Chlortoluron	$C_{10}H_{13}ClN_2O$	212.07196	1.52	60.4	1
22	DEET	$C_{12}H_{17}NO$	191.13123	1.13	98.8	1
23	4-nitrophenol[a,b]	$C_6H_5NO_3$	139.02698	0.24	99.9	2
24	DNOC[a,b]	$C_7H_6N_2O_5$	198.02772	0.26	96.4	2
25	(±)-Abscisic acid[b]	$C_{15}H_{20}O_4$	264.13594	0.83	77.5	2

[a] Identified in both targeted analysis and NTA.

[b] Identified using the IC-HRMS platform.

Appendix 6. Results for ground and drinking water

Appendix 6.1 Ground water – Site E

A suspect screening of five GRUMO pesticide suspects (1,2,4-triazole, desethyl desisopropyl atrazine (DEIA), desisopropyl atrazine (DIA), desphenyl chloridazon (DPC), and methyl desphenyl chloridazon (MDPC), previously identified via targeted analysis as part of the GRUMO analysis in the samples at concentrations 0.01-1.40 µg/L) showed that some target masses of interest, missed by the non-targeted pipeline, could be identified despite not having acquired MS2 data of these. Associated m/z features of DPC could be seen in two samples supposedly containing 0.66 µg/L and 1.40 µg/L (50269 and 51409 respectively) at low intensities (NL < 10,000). An m/z feature corresponding to MDPC was identified in the samples 50269 and 51409 both containing 0.17 µg/L though these features were again low in intensity and lacked a corresponding fragmentation spectrum (MS2). This would indicate a weakness in the applied top20 data-dependent acquisition where relevant low-intensity features would not necessarily be fragmented.

Even when using suspect screening, it was not possible to detect either 1,2,4-triazole, DEIA, or DIA in any of the samples measured by both LC-HRMS and IC-HRMS. Whether this is due to concentrations below the detection limits of the instrument or an inability to resolve these compounds in the two platforms is currently inconclusive, as neither of these compounds were present in the used pesticide standard solution. Both desethyl-atrazine and atrazine, present in the pesticide standard solution, were undetected in IC-HRMS and it is likely that both DEIA and DIA would be undetected as well. Of the five compounds of interest only DPC was present – and detected – in the pesticide standard solution (Figure 15).

It is likely the automated NTA identification workflow could annotate DPC and MDPC in the samples if sufficient MS2-data was available, which could be done by performing iterative data acquisition and/or by implementation of an inclusion list to enable fragmentation of the detected features. This would be highly relevant to test in future measurements. With an ongoing extension of the online (mzCloud) and offline libraries, the feature annotations in the NTA workflow becomes more comprehensive and extensive.

To reliably obtain MS2 data and to improve the annotation confidence in the NTA workflow, the spectra should be recorded using iterative data-dependent acquisition to ensure fragmentation of otherwise ignored low-intensity entities. This acquisition process could potentially become automated in a similar way as described by others[8].

Compound	Formula	Class	Samples							
			50269		50436		51251	51409	51461	51463
			Glass	Plastic	Glass	Plastic	Glass	Glass	Plastic	Plastic
1,2,4-Benzenetricarboxylic acid	C9 H6 O6									
11β-Hydroxyandrosterone	C19 H30 O3	Steroid								
12-Hydroxydodecanoic acid	C12 H24 O3	Fatty acid								
16-Hydroxyhexadecanoic acid	C16 H32 O3	Fatty acid								
1-Tetradecylamine	C14 H31 N									
2,6-Di-tert-butyl-1,4-benzoquinone	C14 H20 O2									
2-Chlorobenzoic acid	C7 H5 Cl O2	Pesticide Metabolite								
2-Hydroxycaproic acid	C6 H12 O3									
2-Hydroxyvaleric acid	C5 H10 O3									
2-Isopropylmalic acid	C7 H12 O5									
3-Phenoxybenzoic acid	C13 H10 O3	Pesticide Metabolite								
4-(hydroxymethyl)benzoic acid	C8 H8 O3									
4-Anisic acid	C8 H8 O3									
4-Hydroxybenzaldehyde	C7 H6 O2									
4-Methylbenzophenone	C14 H12 O									
6-(2-Hydroxypropan-2-yl)-4,8a-dimethyl-2,3,4,4	C15 H28 O3									
9-Oxo-10(E),12(E)-octadecadienoic acid	C18 H30 O3									
Bis(2-ethylhexyl) amine	C16 H35 N									
Caffeic acid	C9 H8 O4	Plant metabolite								
Cetrimonium	C19 H41 N	Topical Antiseptic								
cis-12-Octadecenoic acid methyl ester	C19 H36 O2									
Citraconic acid	C5 H6 O4									
Cyanuric acid	C3 H3 N3 O3									
D-(-)-Ribose	C5 H10 O5	Carbohydrate								
D(+)-Phenyllactic acid	C9 H10 O3									
Decanamide	C10 H21 N O									
Decanoic acid	C10 H20 O2	Fatty acid								
Dibutyl sebacate	C18 H34 O4									
Diethyleneglycol dibenzoate	C18 H18 O5									
Diisobutyl succinate	C12 H22 O4									
Dipropyleneglycol dibenzoate	C20 H22 O5									
(E)-7-Methylidene-10-oxo-4-(propan-2-yl)undec-	C15 H24 O3									
Ethyl myristate	C16 H32 O2	Fatty acid								
Lauric acid	C12 H24 O2									
Methyl dihydrojasmonate	C13 H22 O3									
Mono(2-ethylhexyl) phthalate (MEHP)	C16 H22 O4	Phthalate								
N-Acetylneuraminic acid	C11 H19 N O9									
Nonanoic acid	C9 H18 O2	Fatty acid								
Oleamide	C18 H35 N O	Plant metabolite								
OPEO(Octoxynol)	C16 H26 O2	Polymer								
Salicylic acid	C7 H6 O3									
Tetradecanedioic acid	C14 H26 O4									
Tris(2-ethylhexyl) phosphate	C24 H51 O4 P	Flame retardant								
α-Linolenic acid	C18 H30 O2									
α-Piperidinobutiophenone	C15 H21 N O									
β-Amyrenonol	C30 H48 O2									
δ-Gluconic acid δ-lactone	C6 H10 O6									

Figure 14. Overview of the 47 detected substances (level 2) in samples from site E (id no. 50269, …, 51463). Red indicates a positive identification. Green indicates the molecules is not detected in the sample. Samples were delivered in glass and/or plastic containers. No investigation in the difference between these container materials were pursued in the current study.

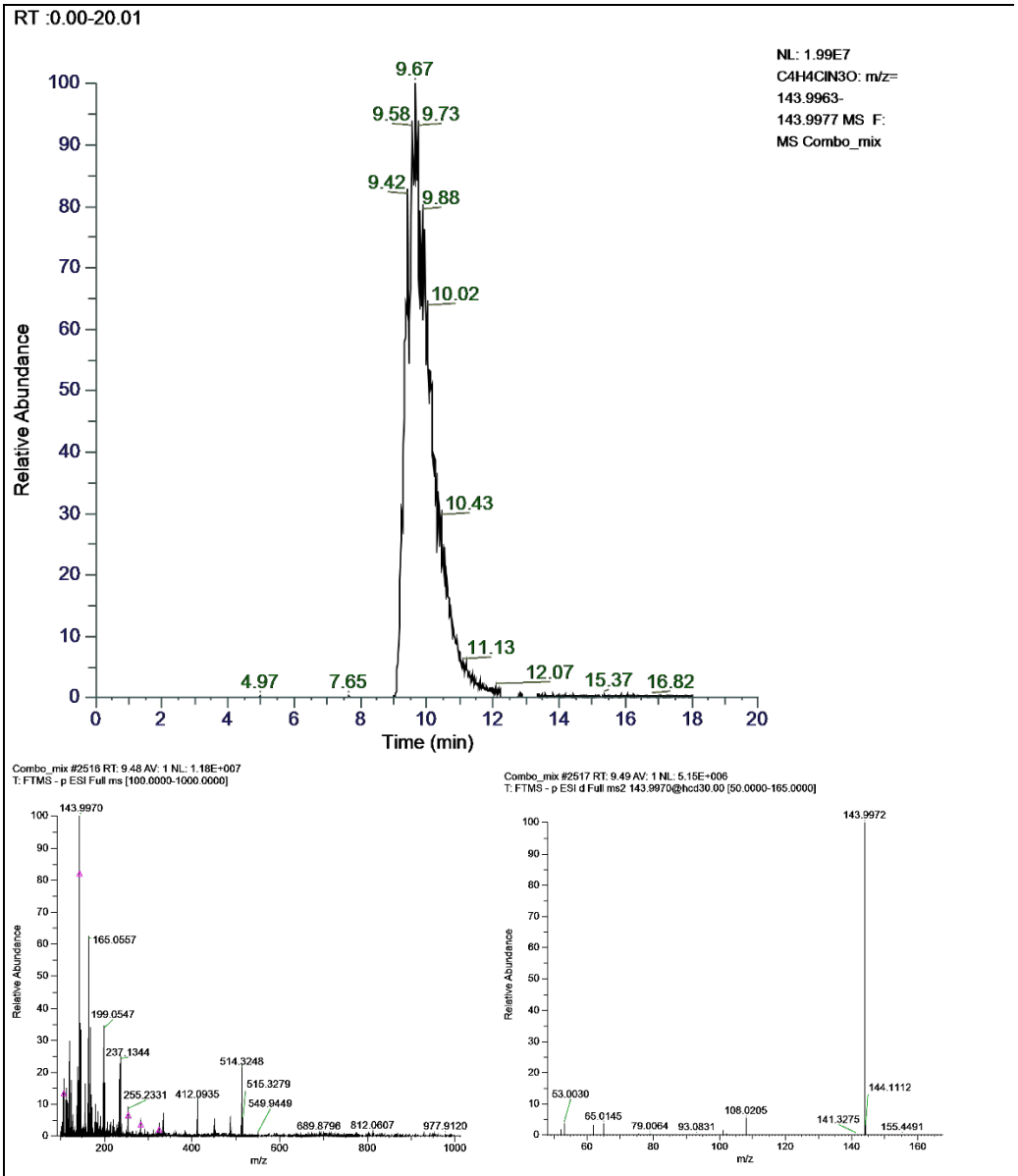

Figure 15. Recorded extracted ion chromatogram of desphenyl chloridazon (DPC) reference standard (top). MS1-spectrum (bottom-left) and MS2-spectrum of m/z 143.9970 (bottom-right).

Appendix 6.2 Drinking water

Waterworks — Site F. Due to evidence of pesticides in the F-samples, a spiking experiment was carried out on the triplicated CLAM-samples and a single field blank, and triplicated grab samples and a single field blank. This was done in order to confirm that the identification was not caused by false positives (Figure 16).

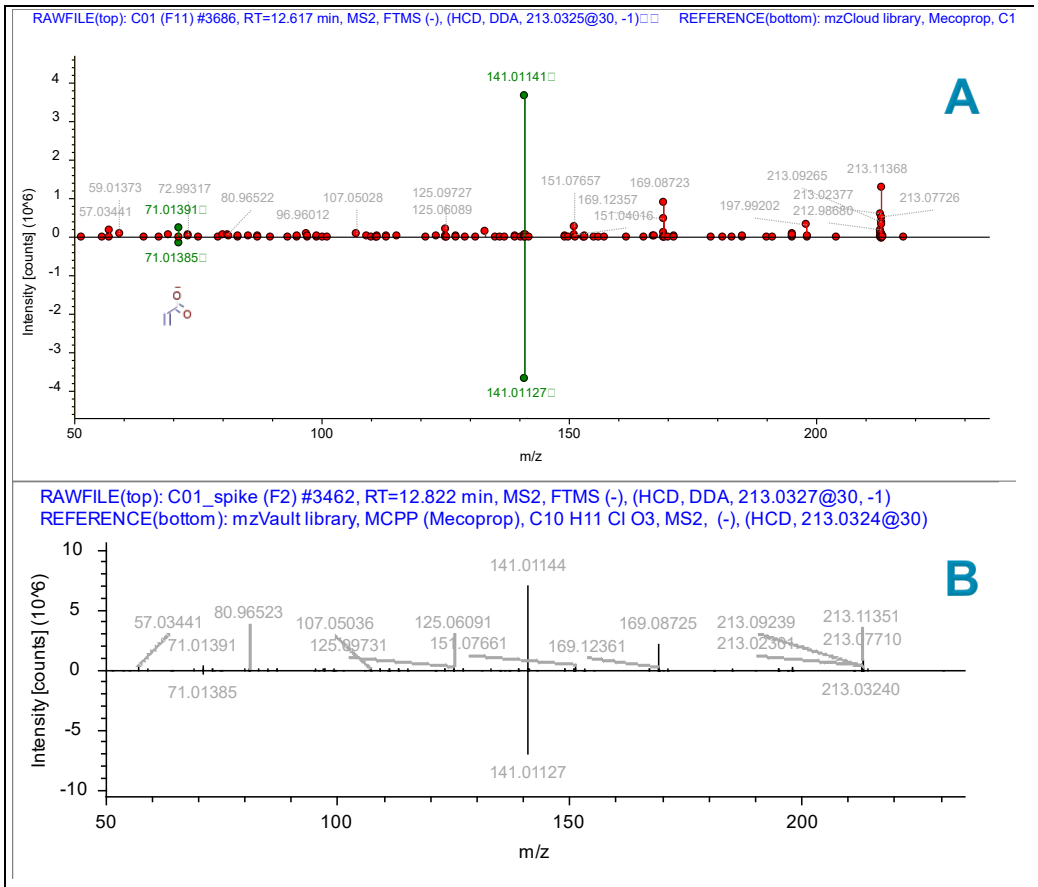

Figure 16. (A) Comparison of (top) measured MS2 spectrum of mecoprop (at m/z = 213.0325) in CLAM samples measured by IC-HRMS with (bottom) spectral library (mzCloud) of mecoprop with a spectral match of 64.1%. (B) Comparison of (top) measured MS2 spectrum of mecoprop (at m/z = 213.0325) spiked in CLAM samples measured by IC-HRMS with (bottom) spectral library (mzCloud) of mecoprop with a spectral match of 79.0%.

In the spiked samples both the MS1 and MS2-data were well resolved showing the MS-spectra for dimethachlor ESA in a blank, non-spiked, and spiked CLAM-sample (Figure 17). Both retention times and m/z-values are the same between the spiked and non-spiked samples and MS1 and MS2 fits as well: consequently, dimethachlor ESA is present at site F. Figure 18 displays the MS1 and MS2 spectrum of spiked dimethachlor ESA, clearly showing identical features.

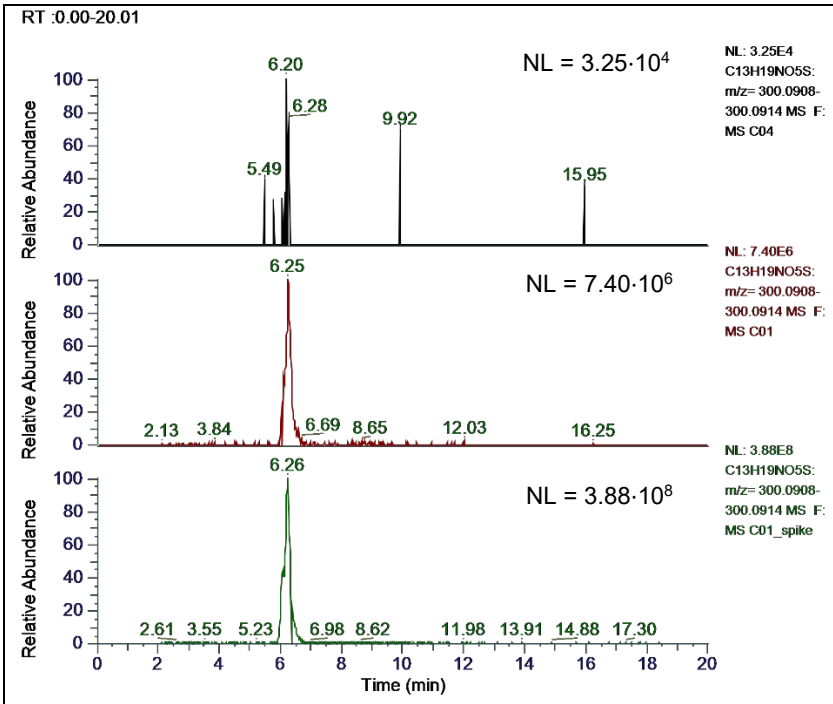

Figure 17: Extracted ion chromatograms for dimethachlor ESA ([M-H]$^-$ @ m/z = 300.0911 in a blank (top), non-spiked (middle), and spiked (bottom) CLAM-sample. Mass tolerance = 1 ppm. Note that the blank signal is 200 times lower than the signal of the non-spiked sample – and 10,000 times lower than the spiked sample signal.

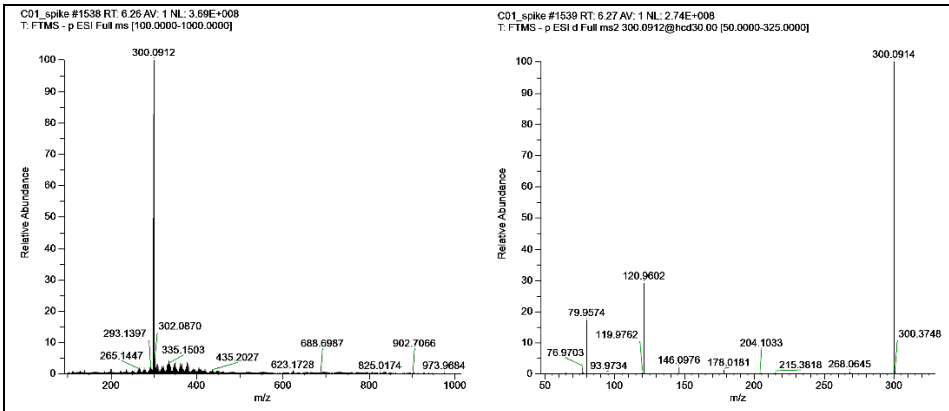

Figure 18. MS1 spectrum for dimethachlor ESA in the spiked CLAM-sample (left). MS2 spectrum of dimethachlor ESA (m/z = 300.0912) in the spiked sample. CE = 30 eV (right).

Though no full quantification could be done, the peak areas between the spiked and non-spiked samples reveal an estimated concentration of dimethachlor ESA in the water samples of 1.8 ± 0.3 ng/L when accounting for sample volume (80 L) and assuming 100% recovery[c]. From this number can see that even an extraction efficiency of 2% would allow for detection at the limit of 0.1 µg/L.

An *in situ* recovery experiment should be initiated in the future to get a more accurate picture of the quantification capabilities of the NTA approach.

[c] $c = \frac{x - x_{\text{blank}}}{x_{\text{spike}} - x} \cdot c_{\text{spike}} \cdot \frac{2V_{\text{vial}}}{V_{\text{sampled}}}$, Where c is the estimated concentration (mg/L) of analyte, x is the signal of the target analyte in the unspiked sample, x_{blank} is the signal of the target analyte in the instrumental blank, x_{spike} is the signal of analyte in the spiked sample, c_{spike} is the added concentration (5.0 mg/L) of analyte in the spiked sample, V_{vial} is the final sample volume after extraction (1 mL), and V_{sampled} is the total volume of extracted water by the CLAM.

Waterworks — Site G. The results from the spiking experiment showed promising results regarding compound identification of anionic pesticides in IC-HRMS. Of the 48 IC in-house library entries, a total of 25 could be fully resolved at either confidence level 1 (13 full matches in both mzCloud and in-house library) or 2 (12 full matches only in the in-house library) in the spiked samples by the established NTA workflow, see Table 2. None of the non-spiked samples indicated the presence of pesticides.

Table 2: Identified pesticides in CLAM and grab samples from site G spiked with 0.1 - 100 µg/L pesticide standard solution analysed by IC-HRMS. These values are effectively up to 100,000 times lower when enriched with CLAM-samples.

Pesticide	**Lower concentration limit (µg/L)**			
Sample matrix	CLAM		Grab	
Spiked	No	Yes	No	Yes
Glyphosate	-	0.1	-	0.1
MCPA	-	0.1	-	0.1
Mecoprop	-	0.1	-	0.1
2-AEP	-	1	-	0.1
2,4-D	-	10	-	10
Cloprop	-	10	-	10
Dichlorprop	-	10	-	10
Picloram	-	10	-	10
Aminopyralid	-	100	-	10
Bentazone	-	100	-	10
Chloramben	-	100	-	10
Tepraloxydim	-	100	-	10
Acephate	-	100	-	100
Chloramphenicol	-	100	-	100
Cinosulfuron	-	100	-	100
Clethodim	-	100	-	100
Clopyralid	-	100	-	100
Cycloxydim	-	100	-	100
Dimethachlor ESA	-	100	-	100
Flumetsulam	-	100	-	100
Fluroxypyr	-	100	-	100
Tralkoxydim	-	100	-	100
Chlorsulfuron	-	-	-	100
Gibberellic Acid	-	-	-	100
Maleic Hydrazide	-	-	-	100

Linear behaviour of the peak areas from the spiking experiments can be seen for glyphosate in the triplicate CLAM-samples, shown for the concentration range of 0.1 – 100 µg/L in Figure 19. Similar linear behaviour is observed for 2-AEP, MCPA, and mecoprop – as well as for the remaining compounds, though this display of linearity is based only upon one or two points.

By running a pure NTA workflow these 25 pesticides were identified at instrumental concentrations of 0.1 - 100 µg/L. Giving that samples are preconcentrated over 100,000 times with the CLAM system, these concentrations could effectively be 100,000 times lower, i.e. potentially having an NTA workflow capable of confidently identifying compounds present in water in concentrations of 0.001 – 1 ng/L – as long as these are retained in the CLAM SPE filter.

The indicated lower concentration limits are effectively up to 100,000 times lower when enriched with CLAM-samples, as these samplers are shown capable of directly extracting >100 L of water easily eluted as 1 mL.

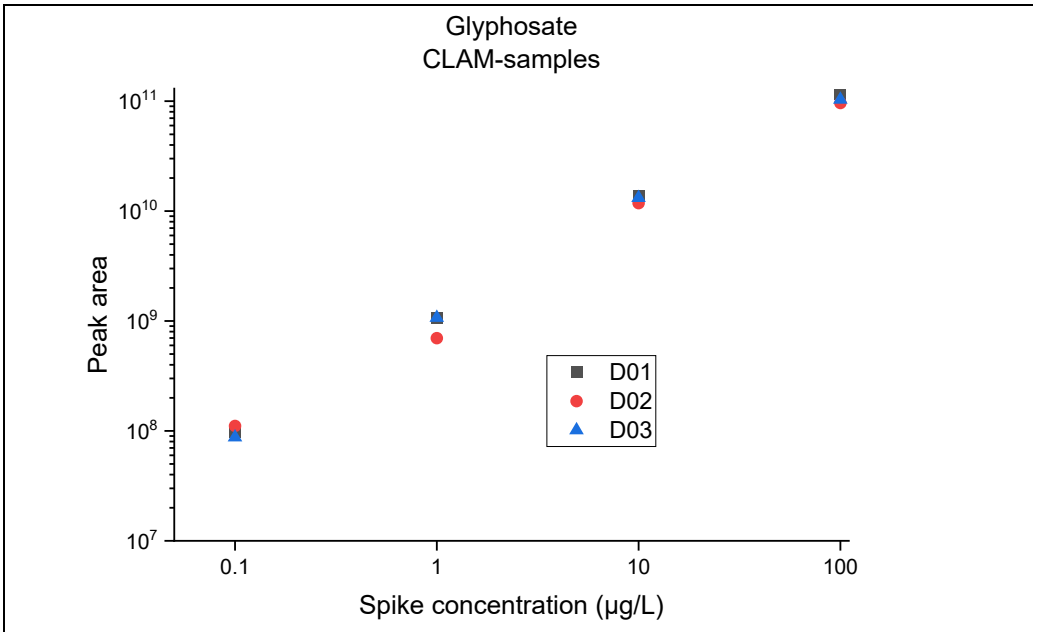

Figure 19. Standard addition of glyphosate in CLAM-triplicates collected from a water work. D01-D03 are the sample ID's of the three CLAM-sample triplicates.

Aminopyralid, bentazone, chloramben, chlorsulfuron, gibberellic acid, maleic hydrazide, and tepraloxydim displayed lower detection limits in the grab samples than the CLAM-samples, suggesting that either certain compounds of interest are sensitive towards matrix effects – ideally being identified in a 'cleaner' matrix such as the grab sample, where the background is expected to be lower than in the CLAM samples – or they are simply not effectively retained in the CLAM HLB SPE filter.

It is likely that adjustments to the post-processing pipeline could increase the instrumental detection limits of the compounds further (for example 'go deeper into the data' by decreasing the peak intensity threshold in the pipeline). For the remaining unidentified compounds from the pesticide standard solution, it is likely that 1) IC-HRMS is not a suitable method for these targets with the current settings and that the targets are better resolved using either LC-HRMS or GC-HRMS, 2) data acquisition was insufficient and could be improved by running iterative data dependent acquisition in addition to an inclusion list (suspect list in Appendix 2), and 3) the compound identification should be complemented by a targeted identification workflow enabling lower detection limits.

Both separation methods (LC and IC) should complement each other for widest range of compound detection. IC-HRMS analysis was sufficient for verification of the protocol. LC-HRMS could have been pursued as well to back-up the claims from the IC-HRMS platform, but was due to time constraints not performed.

HITLIST - Holistic non-targeted approach to determine pesticide and biocide residues in the aquatic environment

Recent technological developments in novel analytical high-resolution mass spectrometry equipment and advanced data processing tools have made a ‘non-targeted analysis’ concept feasible to find known and unknown environmental pollutants.

A momentous drawback of the current applied targeted chemical analysis approaches used in national environmental monitoring programmes is the exclusive focus on a predefined list of compounds for detection. Hence, other chemical entities, potentially also present in the given sample, are not observed and their presence are unnoticed. Non-target analysis is chemical sample analysis without any prior knowledge about its chemical content. The idea is to cover as many chemicals as possible, without focusing on a predefined selection.

This research project HITLIST demonstrates that non-target analysis can elucidate a wide range of pesticides and biocides, as well as other xenobiotics and natural substances, in a broad range of water samples. The project developed suspect and non-targeted screening approaches on commercially available technologies and solutions, enabling other research laboratories, enterprises and academia to establish such non-targeted analysis methodologies. The performed research optimized two high-resolution mass spectrometry platforms hyphenated either with liquid chromatography or ion exchange chromatography. Water samples from various sources, i.e. waterworks, groundwater wells, surface water, coastal water, wastewater effluent and rainwater were analysed by non-targeted analysis. Water samples were prepared for analysis by solid-phase extraction or direct injection. A total of 45 samples were analysed with one or both high-resolution mass spectrometry platforms within the project and in general the non-targeted analysis revealed more than a thousand substances in every sample.

Environmental
Protection Agency
Tolderlundsvej 5
DK-5000 Odense C

www.mst.dk