

# **Environmental Forensics of Industrial Wastewater based on Non-Target Screening**

## **Dissertation**

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

– Dr. rer. nat. –

vorgelegt von

**Kirsten Purschke**

geboren in Osnabrück

Fakultät für Chemie  
der  
Universität Duisburg-Essen

2020



Die vorliegende Arbeit wurde im Zeitraum von Oktober 2017 bis Juni 2020 im Arbeitskreis von Prof. Dr. Torsten C. Schmidt am Institut für Instrumentelle Analytische Chemie der Fakultät für Chemie der Universität Duisburg-Essen durchgeführt.

Tag der Disputation: 18. September 2020

Gutachter: Prof. Dr. Torsten C. Schmidt

Prof. Dr. Oliver J. Schmitz

Vorsitzender: Prof. Dr. Alexander J. Probst

# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub | universitäts  
bibliothek

Diese Dissertation wird über DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

**DOI:** 10.17185/duepublico/72815

**URN:** urn:nbn:de:hbz:464-20200930-091122-1

Alle Rechte vorbehalten.



### Summary

The discharge of industrial wastewater into surface water requires constant monitoring of the water quality to comply with specifications and quality requirements. Furthermore, it must be ensured that individual approvals for the production plants are obtained when industrial wastewater is discharged into the wastewater treatment plant (WWTP). For this reason, the influent of the WWTP is continuously monitored for trace organic compounds (TrOCs). These compounds are usually determined using liquid- (LC) or gas chromatographic (GC) target methods coupled to low-resolution mass spectrometers. However, these methods monitor only a small part of TrOCs in wastewater. Only a limited number of compounds can be detected in a single run and many compounds are ignored in the analysis as they are not part of the target list. Thus, unknown TrOCs neither can be detected nor identified in wastewater samples using these methods, even if they are present in high concentrations. Therefore, high-resolution mass spectrometers (HRMS) have become more and more common in water analysis to carry out more extensive monitoring by detecting both known and unknown compounds. Besides, in combination with a non-target screening (NTS), the generated HRMS data additionally enable the identification of unknown compounds.

The application of LC-HRMS in NTS related to industrial wastewater data is described in this work. Sample treatment procedures and an analytical LC-HRMS method are developed which enable the sensitive and reliable monitoring of TrOCs in a broad polarity range. Additionally, the development of a reliable data processing algorithm for NTS is part of this work. A large amount of data is produced in LC-HRMS that cannot completely be evaluated. Thus, prioritisation methods are required, enabling data reduction. As a result, three prioritisation strategies were developed, which make it possible to extract relevant features (a combination of a particular mass-to-charge ratio, the associated retention time and intensity) from the data for identification. The relevance of each feature depended on the prioritisation strategy.

The first prioritisation method selects these features, whose intensities followed rising or falling trends over time-series measurements. As a result, influences on industrial wastewater through the different production processes in an industrial park were recognised. This method was carried out by principal component analysis (PCA) and group-wise PCA (GPCA). 130 of initially 3303 detected features were prioritised in the WWTP influent samples. In addition to prioritisation, the introduced method enabled componentisation (grouping of several features into one TrOC). As proof-of-concept, one feature with an increasing trend over five months was identified as N-methylpyrrolidone.

In the second trend-related prioritisation method, the time series investigations were linked to spatial trends. For this purpose, several sampling sites before and after the WWTP were sampled and analysed over five months. This allows evaluating the treatment procedure of industrial wastewater over time. Besides, site-specific features were detected. In future studies, these features could serve as a fingerprint in the monitoring of the wastewater streams.

In contrast to the first two, the latter prioritisation method shows a more technical approach. TrOCs, which were repeatedly detected in the influent of the WWTP by the routine monitoring, but which were initially not identified (*'known unknowns'*) are prioritised and identified by an (offline) two-dimensional LC coupled to two kinds of detection techniques (ultra-violet detection and MS). The identification of these *'known unknowns'* is of high interest for the operators of the WWTP. LC-UV peaks from wastewater samples were fractionated manually in the first dimension and elucidated in the second dimension, the LC-HRMS, by NTS. By applying this method, the analysis of only one sample fraction led to sampling purification and therefore, data reduction. As an example, the *'known unknown'* with the retention time of 41.1 minutes and the maximum UV absorption of 240 nm in the first dimension was successfully identified as a dichlorodinitrophenol isomer.

All in all, this work shows that the use of HRMS data, in combination with NTS and the application of the presented prioritisation methods, extends the monitoring of industrial wastewater and permits an evaluation of the WWTP processes. It indicates great potential for future establishment in routine monitoring.

### Zusammenfassung

Die Einleitung von Industrieabwässern in Oberflächengewässer erfordert eine ständige Überwachung der Wasserqualität zur Einhaltung von Spezifikationen und Qualitätsanforderungen. Zusätzlich muss sichergestellt werden, dass Einzelgenehmigungen der Produktionsanlagen bei der Einleitung von Industrieabwässern in die Kläranlage eingehalten werden. Daher werden die Abwasserproben hinsichtlich von Spuren organischer Verbindungen (engl. *trace organic compounds*, TrOCs) permanent überwacht. In der Routineanalytik werden diese Verbindungen üblicherweise mit flüssig- (engl. *liquid chromatography*, LC) oder gaschromatographischen (GC) Systemen aufgetrennt und mit niedrig auflösenden Massenspektrometern detektiert. Diese Analyt gerichtete Vorgehensweise erfasst jedoch nur einen kleinen Teil der TrOCs im Abwasser. Unbekannte TrOCs in Abwasserproben können mit dieser Vorgehensweise weder nachgewiesen noch identifiziert werden, selbst wenn diese in hohen Konzentrationen in der Probe vorliegen. Das liegt daran, dass bei diesen Methoden nur eine begrenzte Anzahl an Verbindungen in einem analytischen Lauf bestimmt werden kann. Außerdem werden viele Verbindungen bei der Analyse ignoriert, da sie nicht Teil der Analyt-Liste sind. Aus diesem Grund werden hochauflösende Massenspektrometer (HRMS) in der Wasseranalyse immer häufiger eingesetzt, um eine umfassendere Überwachung, durch den Nachweis von sowohl bekannten als auch unbekanntem Verbindungen, zu ermöglichen. Zusätzlich ermöglichen die generierten HRMS-Daten in Kombination mit einem Non-Target-Screening (NTS) die Identifizierung dieser unbekanntem Verbindungen.

Diese Arbeit beschäftigt sich mit der Anwendung des NTS Konzeptes auf Industrieabwasserdaten die mittels der LC-HRMS Technik aufgenommen wurden. Es wurden Probenvorbereitungen und eine Analysenmethode entwickelt, die es ermöglichen TrOCs in einem großen Polaritätsbereich empfindlich und verlässlich nachzuweisen. Zusätzlich ist die Entwicklung eines verlässlichen Datenauswertalgorithmus für das NTS Bestandteil dieser Arbeit. Dadurch, dass die Anwendung von HRMS-Daten in Verbindung mit dem NTS, bekannte und unbekanntem Verbindungen detektiert, entsteht eine große Datenmenge, die es zu bewältigen gilt. Um die Aufklärung von unbekanntem Verbindungen realisieren zu können, sind Priorisierungsmethoden erforderlich, die unter anderem zur Datenreduzierung verwendet werden können. Folglich wurden drei Priorisierungsstrategien erarbeitet, die es ermöglichen sollen relevante *Feature* (Zuschnitt von einem bestimmten Masse-zu-Ladungsverhältnis, der zugehörigen Retentionszeit und Intensität) aus der Datenmenge für die Identifizierung zu extrahieren. Die Relevanz richtete sich dabei nach der Priorisierung.

Die erste Priorisierungsmethode selektiert solche *Feature*, deren Intensitäten über einen längeren Messzeitraum, steigenden oder sinkenden Trends folgen. Dadurch sollten Einflüsse auf das Industrieabwasser durch die unterschiedlichen Produktionsprozesse in einem Industriepark erkannt werden. Durchgeführt wurde diese Methode durch die Hauptkomponentenanalyse (engl. *principal component analysis*, PCA) und gruppenweise PCA (GPCA). Von 3303 detektierten *Feature* aus Kläranlagenzuflussproben konnten 130 relevante *Feature* priorisiert werden. Außerdem zeigte sich, dass die Methode zusätzlich zur Priorisierung noch eine Komponentisierung (Gruppierung von mehreren *Features* zu einem TrOC) der detektierten *Feature* ermöglichte. Als konzeptioneller Beweis wurde exemplarisch ein *Feature* mit steigendem Trend als N-Methylpyrrolidon identifiziert.

In der zweiten trendbezogenen Priorisierungsmethode wurden die Zeitreihenuntersuchungen anschließend mit räumlichen Trends verknüpft. Dazu wurden mehrere Probenahmestationen vor und nach der Kläranlage über einen längeren Zeitraum beprobt und analysiert. Dadurch war es möglich das Kläranlagenverhalten von Industrieabwässern über fünf Monate zu bewerten. Außerdem konnten solche *Feature* gefunden werden, welche jeweils für eine Probenahmestation spezifisch waren und somit als Fingerabdruck in der Überwachung der Abwasserströme dienen könnten.

Die letzte Priorisierungsmethode zeigt im Gegensatz zu den ersten Beiden einen eher technischen Ansatz. TrOCs, die in dem Kläranlagenzulauf durch die Routineüberwachung wiederholt detektiert wurden, die zunächst aber nicht eindeutig identifizierbar waren und deren Aufklärung für die Betreiber der Kläranlage folglich im Vordergrund steht („*bekannt* *Unbekannt*“), wurden durch eine (offline) zweidimensionalen LC gekoppelt an zwei unterschiedliche Detektionstechniken (UV und MS) priorisiert und anschließend identifiziert. LC-UV Peaks aus Abwasserproben wurden so in der ersten Dimension manuell fraktioniert und in der zweiten Dimension, der LC-HRMS, durch NTS aufgeklärt. Die Analyse von ausschließlich einer Probenfraktion führt zur Probenreinigung und damit zur Datenreduktion. Exemplarisch wurde die „*bekannt* *Unbekannt*“ mit der Retentionszeit von 41,1 Minuten in der ersten Dimension und der maximale UV-Absorption von 240 nm erfolgreich als ein Isomer der Dichlordinitrophenole identifiziert.

Insgesamt zeigt diese Arbeit, dass die Verwendung von HRMS-Daten, in Kombination mit dem NTS und unter Verwendung der vorgestellten Priorisierungsmethoden, die Überwachung von Industrieabwässern erweitert und eine Bewertung der Kläranlage zulässt. Sie zeigt ein großes Potential für die zukünftige Etablierung in die Routineüberwachung.



---

## Table of Contents

<b>Summary</b> .....	<b>V</b>
<b>Zusammenfassung</b> .....	<b>VII</b>
<b>Table of Contents</b> .....	<b>IX</b>
<b>Chapter 1 Introduction and Theoretical Background</b> .....	<b>- 14 -</b>
1.1 Monitoring of Industrial Wastewater .....	- 14 -
1.2 Basics of Liquid Chromatography coupled to Mass Spectrometry.....	- 15 -
1.3 High-Resolution Mass Spectrometry .....	- 18 -
1.4 Non-Target Screening.....	- 21 -
1.5 Strategies of Non-Target Screening in Industrial Wastewater Analysis .....	- 24 -
1.6 References .....	- 25 -
<b>Chapter 2 Scope and Aims of the Thesis</b> .....	<b>- 32 -</b>
<b>Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater</b> .....	<b>- 35 -</b>
3.1 Introduction.....	- 35 -
3.2 Experimental Section .....	- 37 -
3.2.1 Chemicals and Reagents .....	- 37 -
3.2.2 Sample Materials .....	- 37 -
3.2.3 Quality Control .....	- 38 -
3.2.4 Sample Preparation .....	- 38 -
3.2.5 Sample Storage .....	- 39 -
3.2.6 LC-ESI-qTOF-MS Conditions.....	- 39 -
3.2.7 Data Processing and Data Analysis .....	- 40 -
3.3 Results and Discussion.....	- 42 -
3.3.1 Method Development and Optimisation .....	- 42 -
3.3.1.1 Handling of Sample Preparation .....	- 42 -
3.3.1.2 Stability Analysis during Sample Storage.....	- 43 -
3.3.1.3 Liquid Chromatographic Conditions .....	- 44 -
3.3.1.4 Fragmentation Method in Mass Spectrometry.....	- 46 -
3.3.2 Data Processing .....	- 47 -
3.3.3 Validation of Non-Target Screening based on developed LC-HRMS .....	- 50 -
3.4 Conclusion.....	- 51 -
3.5 References .....	- 52 -
3.6 Chapter Appendix.....	- 56 -
3.6.1 Supplement for Chapter Materials and Method.....	- 56 -

## Table of Contents

---

3.6.1.1	Targets and ISTD .....	- 56 -
3.6.1.2	Non-target screening method.....	- 60 -
3.6.1.3	ISTD stability over Batch.....	- 62 -
3.6.2	Supplement for Chapter Results .....	- 63 -
3.6.2.1	Centrifugation .....	- 63 -
3.6.2.2	Sample Dilution.....	- 64 -
3.6.2.3	Stability Analysis during Sample Storage.....	- 66 -
3.6.2.4	Long-term Storage Conditions .....	- 67 -
3.6.2.5	LC-Column Tests.....	- 67 -
3.6.2.6	Comparison of Fragmentation Methods .....	- 73 -
3.6.2.7	Peak Picking using Intensity Threshold.....	- 74 -
3.6.2.8	Retention Time and Mass Error .....	- 77 -
3.6.2.9	Validation of Non-Target Screening Method .....	- 78 -
3.6.2.10	Application to authentic standards .....	- 79 -
<b>Chapter 4</b>	<b>Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches .....</b>	<b>- 83 -</b>
4.1	Introduction.....	- 83 -
4.2	Experimental Section.....	- 85 -
4.2.1	Solvents and Chemicals .....	- 85 -
4.2.2	Sample Material.....	- 85 -
4.2.3	LC-HRMS Screening Analysis .....	- 86 -
4.2.4	Quality Control.....	- 86 -
4.2.5	Data Treatment and Chemometric Analysis.....	- 87 -
4.2.5.1	Non-Target Data Processing Workflow .....	- 88 -
4.2.5.2	The Fusion of Target Data for Constructing Data Matrix of Both Ionisation Modes .....	- 89 -
4.2.5.3	Multivariate Data Analysis Strategy.....	- 89 -
4.2.5.3.1	Row and Column-wise PCA for Exploratory Analysis.....	- 90 -
4.2.5.3.2	GPCA for Feature Prioritisation.....	- 90 -
4.2.5.4	Identification of Relevant Features.....	- 90 -
4.3	Results and Discussion.....	- 91 -
4.3.1	Multivariate Time Trend Analysis of Target Data.....	- 91 -
4.3.1.1	Column-wise PCA for Prioritisation of Time Trends with High Variation in Intensity of Target Data.....	- 91 -
4.3.1.2	GPCA for Grouping Time Trends of Target Data .....	- 93 -
4.3.1.3	Componentisation of Data Processing Method .....	- 95 -
4.3.2	Application to Industrial WWTP Influent Wastewater Samples.....	- 95 -

## Table of Contents

---

4.3.2.1	Pre-Filtering of Data Matrix using Univariate Statistic .....	- 95 -
4.3.2.2	Prioritisation of Relevant Features in WWTP Influent Samples .....	- 96 -
4.3.2.3	Identification Experiment.....	- 99 -
4.4	Conclusion.....	- 102 -
4.5	References .....	- 103 -
4.6	Chapter Appendix.....	- 106 -
4.6.1	Supplement Information of Chemometric Methods .....	- 106 -
4.6.1.1	PCA for Exploratory Time Trend Assessment.....	- 106 -
4.6.1.2	Group-wise PCA (GPCA).....	- 107 -
4.6.2	Supplement of Experimental Section .....	- 108 -
4.6.2.1	Data sets, Target Compounds and Internal Standards.....	- 108 -
4.6.2.2	LC-HRMS Analysis .....	- 113 -
4.6.2.3	Analytes for Quality Control (QC).....	- 113 -
4.6.3	Supplement of Results.....	- 118 -
4.6.3.1	Results of Training Data .....	- 118 -
4.6.3.1.1	Overview of Detection of ISTD in Training Data Set.....	- 118 -
4.6.3.1.2	Column-wise PCA.....	- 119 -
4.6.3.1.3	Row-wise PCA.....	- 121 -
4.6.3.1.4	Results of GPCA.....	- 122 -
4.6.3.2	Results of Wastewater Samples .....	- 125 -
4.6.3.2.1	Pre-Filtering of Raw Data Matrix .....	- 125 -
4.6.3.2.2	Results of Prioritising Samples in Time Series by Row-wise PCA.....	- 127 -
4.6.3.2.3	GPCA Results and Prioritised Features Responsible for Trend Phenomena using First Prioritisation Route.....	- 129 -
4.6.4	References .....	- 133 -
<b>Chapter 5</b>	<b>Spatial Trend Detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant.....</b>	<b>- 135 -</b>
5.1	Introduction.....	- 135 -
5.2	Experimental Section.....	- 137 -
5.2.1	Chemicals and Reagents.....	- 137 -
5.2.2	Sampling Sites and Storage.....	- 137 -
5.2.3	Trend Detection Workflow.....	- 138 -
5.2.3.1	LC-HRMS Acquisition .....	- 138 -
5.2.3.2	Data Processing .....	- 139 -
5.2.3.3	Assessment of Wastewater Treatment Processes .....	- 140 -
5.2.3.4	Identification Procedures .....	- 142 -
5.2.4	Quality Control.....	- 142 -

## Table of Contents

---

5.3	Results and Discussion.....	- 143 -
5.3.1	Characterisation of Features of studied Sampling Sites.....	- 143 -
5.3.1.1	Comparison of Sampling Sites.....	- 143 -
5.3.1.2	Fingerprinting of Site-Specific Features .....	- 146 -
5.3.2	Assessment of the WWTP using Trend Analysis .....	- 148 -
5.3.2.1	Spatial Trend Analysis .....	- 148 -
5.3.2.2	Temporal Trend Analysis .....	- 151 -
5.4	Conclusion.....	- 154 -
5.5	References .....	- 154 -
5.6	Chapter Appendix.....	- 157 -
5.6.1	Supplement for Chapter Materials and Method.....	- 157 -
5.6.1.1	Used Target Analytes and Internal Standards.....	- 157 -
5.6.1.2	LC-HRMS Method Parameters .....	- 159 -
5.6.1.3	Data Treatment using MarkerView.....	- 160 -
5.6.1.4	Quality Control of ISTD during Data Analysis.....	- 161 -
5.6.2	Supplement for Chapter Results .....	- 162 -
5.6.2.1	Analysis of Sampling Sites.....	- 162 -
5.6.2.2	Spatial Trend Analysis .....	- 165 -
5.6.2.3	Temporal Trend Analysis .....	- 167 -
<b>Chapter 6</b>	<b>Identification of Unknowns in Industrial Wastewater using Offline 2D Chromatography and Non-Target Screening .....</b>	<b>- 172 -</b>
6.1	Introduction.....	- 172 -
6.2	Experimental Section.....	- 173 -
6.2.1	Chemicals and Reagents.....	- 173 -
6.2.2	Sample Material.....	- 174 -
6.2.3	Analytical Methods.....	- 174 -
6.2.3.1	LC Separation and UV Detection in the First Dimension .....	- 174 -
6.2.3.2	LC-HRMS Separation and Detection in the Second Dimension .....	- 175 -
6.2.3.3	<sup>1</sup> H-NMR Analysis for Structure Confirmation.....	- 176 -
6.3	Results and Discussion.....	- 176 -
6.3.1	Fractionation of 'Known Unknown' in the First Dimension.....	- 176 -
6.3.2	Identification Process of prioritised Compound with LC-HRMS Analysis of the Selected LC-UV Fraction .....	- 176 -
6.3.3	Identity Verification of ' <i>Known Unknown</i> ' by <sup>1</sup> H-NMR.....	- 181 -
6.3.4	Quantitative Screening of identified ' <i>Known Unknown</i> ' .....	- 182 -
6.4	Conclusions .....	- 182 -
6.5	References .....	- 183 -

## Table of Contents

---

6.6	Chapter Appendix.....	- 185 -
6.6.1	Supplement for Chapter Materials and Methods .....	- 185 -
6.6.1.1	Instrument Resolution .....	- 185 -
6.6.1.2	Used Analytes .....	- 186 -
6.6.1.3	Different Gradients for Toxicology Determination with HPTLC .....	- 188 -
6.6.1.4	Toxicity test with HPTLC.....	- 190 -
6.6.1.5	Signal Cutting for Manual Fractionation .....	- 191 -
6.6.2	Supplement for Chapter Results and Discussion .....	- 191 -
6.6.2.1	Determination of double bond equivalents .....	- 191 -
6.6.2.2	Constitutional Isomers of Dichlorodinitrophenol .....	- 192 -
6.6.2.3	Toxicity Test with HPTLC.....	- 192 -
6.6.2.4	Results of Quantification Analysis.....	- 196 -
6.6.3	References .....	- 196 -
<b>Chapter 7</b>	<b>General Conclusions and Outlook.....</b>	<b>- 197 -</b>
7.1	References .....	- 201 -
<b>Chapter 8</b>	<b>Appendix.....</b>	<b>- 204 -</b>
8.1	List of Abbreviations .....	- 204 -
8.2	List of Figures and Figures-A.....	- 210 -
8.3	List of Tables and Tables-A .....	- 215 -
8.4	List of Publications.....	- 220 -
8.4.1	Articles in Peer-Reviewed Journals.....	- 220 -
8.4.2	Other Articles.....	- 220 -
8.4.3	Oral Presentations .....	- 220 -
8.4.4	Poster presentations.....	- 220 -
8.5	Declaration of Scientific Contribution .....	- 222 -
8.6	Curriculum Vitae .....	- 223 -
8.7	Erklärung .....	- 224 -
8.8	Danksagung .....	- 225 -

## Chapter 1 Introduction and Theoretical Background

### 1.1 Monitoring of Industrial Wastewater

Industrial wastewater is a by-product of industrial processes. Whether it is food, clothing, paper or chemical products, water is required for nearly every step of production. The acidic or alkaline wastewater from the production plants, laboratories and technical equipment needs to be treated complying with environmental protection laws. It may contain suspended and dissolved organic matter, inorganic salts, pathogens and nutrients occurred as starting materials, by- or waste-products resulting from production (Dvořák et al., 2014). Its diversity is further increased by chemical reactions in the wastewater streams and by transformation reactions in the wastewater treatment plant (WWTP). This leads to a greater analytical challenge compared with monitoring of the less varied municipal wastewater (Gupta and Bux, 2019, p. 89).

In the WWTP, wastewaters are treated to improve the water quality to a level that allows its discharge back into the environment (Eggen et al., 2014). Before the wastewater flows into the WWTP, comprehensive analysis is performed. The purpose is to ensure that the WWTP will be able to treat the current compound load properly. Substances which cannot be eliminated biologically, or interfere with a biological purification, are already removed before the sewage system to protect the WWTP treatment step and hence the environment of undesirable discharge of trace organic compounds (TrOCs). To that end, a storage reservoir serves as a buffering capacity between the industrial sewer system and the inlet of the WWTP. Within the flow between the sewer of the production site and the WWTP, the water is analysed. For each production plant, specifications or individual authorisations declare, which substances may be discharged in which quantities and during which period into the WWTP. To check compliance with these legal processes, substantial monitoring of industrial wastewater is essential. Therefore, analytical methods that are robust, safe and sensitive are needed. Routinely, monitoring is performed with gas chromatography coupled to flame ionisation detection (GC-FID) (Oyetibo et al., 2017), liquid chromatography coupled to ultra-violet detection (LC-UV) (Purschke et al., 2020), ion chromatography (IC) methods (Özkaraova et al., 2018) and inductively coupled plasma optical emission spectrometry (ICP-OES) or mass spectrometry (ICP-MS) analysis (Özkaraova et al., 2018), as well as several online analysis techniques, e.g. determination of total organic carbon (TOC), temperature or pH measurements (Kessler, 2006). For sensitive and selective determination of TrOCs, LC- and GC-methods coupled to low-resolution mass spectrometry (LRMS) detectors (Kumari and Tripathi, 2019; Wortberg and Kurz, 2019) are increasingly used. However, compounds are still

measured by prior compound-specific tuning. Numerous parameters must be optimised for each analyte and the number of compounds, which can be analysed, is limited (Van der Heeft et al., 2009). Therefore, unknown substances neither can be detected nor identified in wastewater samples, even if they are present in high concentrations. Consequently, the demand for analytical methods going beyond targeted analytes, widening the analytical window, increased in recent years.

### 1.2 Basics of Liquid Chromatography coupled to Mass Spectrometry

For monitoring organic compounds in wastewater, chromatographic separation processes with selective detection are required. The coupling of liquid chromatography with mass spectrometry (LC-MS, see Figure 1-1) is a suitable technique for separation, detection and identification of organic molecules in trace level concentration, i.e.  $\mu\text{g L}^{-1}$  and lower.

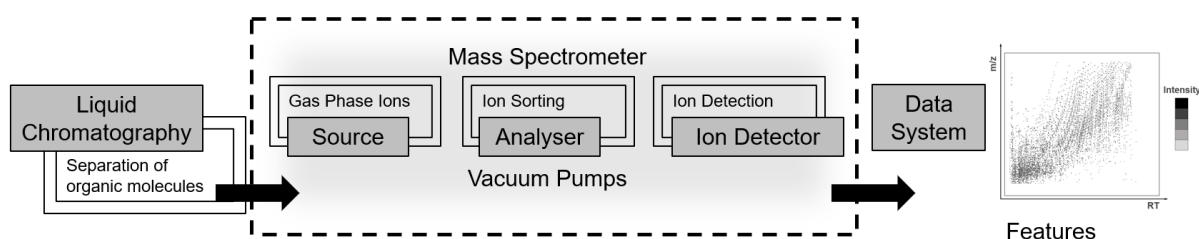


Figure 1-1 – Components of LC-MS. The LC separates organic molecules. In the ionisation source ions are produced, which are separated according to their mass-to-charge ratio ( $m/z$ ) in the mass analyser. The detector collects the separated ions and converts them into electrical signals. Finally, the data system processes the signals resulting in three-dimensional data of  $m/z$ , retention time and intensity (features (Bader et al., 2016)).

High-performance liquid chromatography (HPLC) is used for the determination of thermally labile or low-volatility, (slightly) polar compounds (Krauss et al., 2010). The separation is based on the different distribution of each analyte between the stationary (column) and the mobile phase (eluent). The injected sample aliquot is pumped together with the mobile phase under high pressure through the separation column. The stronger the interaction with the stationary phase, the longer the analyte interacts with column material, increasing analyte's retention. Mainly two types of stationary phase materials are available, normal phase (NP) and reversed-phase (RP). In RP chromatography, alkyl chains (mostly C18) chemically bonded on silica are used as the stationary phase. The stationary phases are characterised by their hydrophobic interactions and are, therefore, particularly suitable for non-polar compounds, due to their stronger interaction. Analytes are eluted in decreasing polarity order. However, C18 stationary phases could be used for screening methods of slightly polar compounds and non-polar compounds. Since not all silanol groups of stationary phase material can be chemically converted during the production process, because of steric hindrance, interactions

between the remaining free silanol groups and polar groups of the analyte compounds can occur (Thelen and Rösen, 2017). As a disadvantage, depending on whether the silanol groups are protonated or deprotonated, H-donor/-acceptor analytes can interact with them supporting peak broadening (Neue et al., 2001). By end-capping, remaining silanol groups can be partially derivatised using, e.g. trimethylchlorosilane. Therefore, less interaction of basic analyte residuals is possible, which results in more symmetrical peaks. Furthermore, polar embedded RP stationary phases can be used for selective retention of compounds (O'Gara et al., 1999). These phases contain functional groups, such as amides or carbamates, embedded in the alkyl chains of RP columns. In water analysis, both end-capped and polar embedded RP stationary phases are used (Bader et al., 2016; Blum et al., 2017; Schymanski et al., 2014b). NP is less commonly used in water analysis because the handling of aqueous samples would require sample preparations (Nagy and Vékey, 2008). However, in recent years, hydrophilic interaction liquid chromatography (HILIC) that has similar separation properties to NP chromatography has become a new trend in water analysis. HILIC enables the separation of highly polar compounds in water by establishing a hydrophilic environment of the stationary phase (Rüdel et al., 2020).

The choice of mobile phase depends on the polarity and the associated elution strength, which is needed to elute the analytes (Gritter et al., 1987; Meyer, 2009). Complex matrices require selective separation to reduce matrix effects on ionisation, to minimise background interferences and to prevent spectral complexity. Matrix effects can lead to signal suppression or signal enhancement of analytes, which affects the sensitivity, leading to errors in quantitative analysis and hinder data interpretation (Bader et al., 2017; Fang et al., 2015; Schollée et al., 2016). These effects depend on the chromatographic retention time (Trufelli et al., 2011) and are usually higher on the early-eluting peaks, but should not be excluded for later retention times (Matuszewski et al., 2003).

The interface between liquid chromatography and mass spectrometer is the ion source (see Figure 1-1). The analytes are ionised and transferred from liquid to gaseous phase while completely removing the solvent. The ionisation methods of LC-MS are mainly atmospheric pressure ionisation (API), such as the electrospray ionisation (ESI) and atmospheric pressure chemical ionisation (APCI). LC-ESI-MS is one of the state-of-the-art techniques in water analysis (Freeling et al., 2019; Leendert et al., 2015; Rager et al., 2016; Robles-Molina et al., 2014). Using ESI, by continuously charging and spraying of the LC effluent, ions are generated inside the ESI source and transported into the gaseous phase (Rohner et al., 2004). The fluid sample is passed through a steel or quartz capillary. The droplet formation takes place with the aid of nitrogen in the strong electric field, whereby a liquid cone (Taylor cone) is formed. At the end of the cone, small positively or negatively charged droplets are formed. The aerosol is



dried, reducing the droplet radius until the charge density and the surface tension becomes critical and the droplet bursts due to electrostatic repulsion (Coulomb explosion). ESI can be performed in either positive or negative mode (Holčapek et al., 2010). The positive mode is more likely to be used for basic molecules such as nitrogen compounds, ethers, esters, thiols and epoxides. It leads to positive ionisation and increases the mass of the molecular ion by one Dalton (Da) for the most frequently observed protonated molecule  $[M+H]^+$ . Negative ionisation reduces the mass of the molecular ion by one Da for the most frequent deprotonated molecule  $[M-H]^-$ , which is mainly used for acidic substances such as carboxylic acids, phenols, phosphates and nitrogen oxides. Since ESI is a so-called soft ionisation method, it is suitable for the ionisation of thermolabile, polar and non-volatile substances. However, it can still lead to fragment ions and adducts in the source additionally to the (de-)protonated molecules  $[M+H]^+/[M-H]^-$  that complicate the interpretation of data (Keller et al., 2008). Some significant adducts and fragments occurring in ESI are listed in Table 1-1.

Table 1-1 – Adducts and fragments typically occurring in ESI ionisation (Keller et al., 2008).

	<b>ESI (+)</b>	<b>ESI (-)</b>
Adducts	$[M+H]^+$ , $[M+Na]^+$ , $[M+NH_4]^+$ , $[M+K]^+$ , $[M+nH]^{n+}$	$[M-H]^-$ , $[M+HCOO]^-$ , $[M+Cl]^-$
Fragments	$[M+H-H_2O]^+$ , $[M+H-CO_2]^+$ , $[M+H-C_2H_6O]^+$	$[M-H-CO_2]^-$ , $[M-F]^-$

After ionisation, the ions are separated according to their mass-to-charge ratio ( $m/z$ ) in the mass analyser (see Figure 1-1) of the mass spectrometer. In literature, further terms instead of  $m/z$  are used. The 'exact mass' is the theoretical mass obtained by calculating the mass of the molecule or ion with a specified isotopic composition (isotopes  $\triangleq$  atoms of the same element that have different numbers of neutrons in their nuclei) and the 'accurate mass' is the experimentally detected mass (cf. Figure 1-2). The 'nominal mass' of a molecule or an ion is the sum of the most abundant naturally occurring stable isotopes of the elements (Murray et al., 2013). The mass accuracy can be assigned as the absolute mass error (accurate mass minus exact mass, typically in mDa) or as the relative mass error in ppm (absolute mass error divided by exact mass and multiplied with  $10^6$ ).

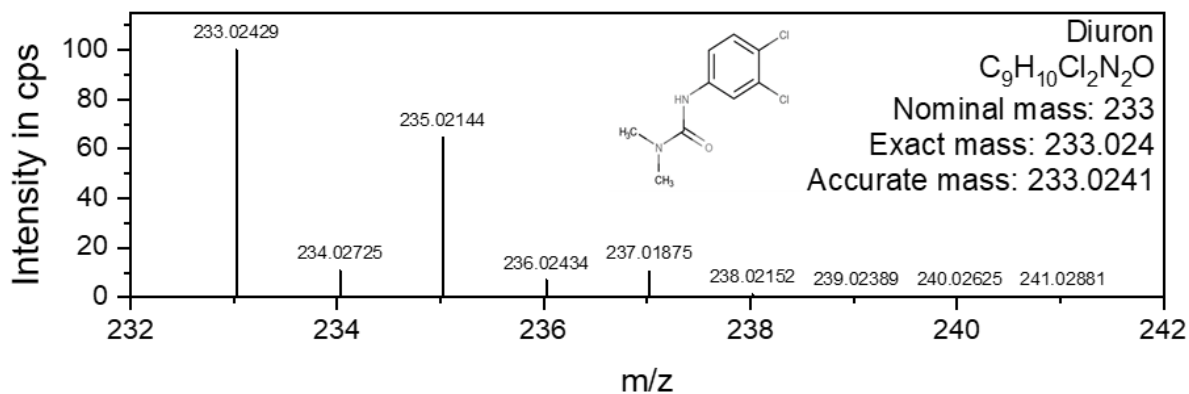


Figure 1-2 – Illustration of mass spectrometry terms on the example of the herbicide diuron. Listed  $m/z$  values correspond to the protonated molecule  $[M+H]^+$ . The mass spectrum shows the distinct isotopic pattern of two chlorine atoms of the structure:  $C_9H_{10}^{35}Cl_2N_2O \triangleq$  nominal mass of 233 (100%),  $C_9H_{10}Cl^{35}Cl^{37}N_2O \triangleq$  nominal mass of 235 (64.71%),  $C_9H_{10}Cl^{37}Cl^{35}N_2O \triangleq$  nominal mass of 237 (10.70%). The other occurring ions describe the  $^{13}C$ -isotopes (1% of  $^{12}C$  isotopes).

For the LC-MS technique, commonly quadrupoles (Q), ion traps, time-of-flight (TOF) or Orbitrap mass analysers are used (Crimmins and Holsen, 2019; Haag et al., 2016). To improve the selectivity and sensitivity of the mass spectrometers, several analysers (see Figure 1-1) are equipped with an additional fragmentation and separation step (LC-MS/MS). The first MS is used as a filter to select the precursor ion, followed by fragmentation with high energy and inert gas, e.g. nitrogen. The second mass analyser is responsible for the separation of the product ions, generated by the fragmentation (Bristow, 2006; Mellon et al., 2002). This is usually done in triple quadrupole (QQQ) analysers and quadrupole-time-of-flight (qTOF) devices. The advantages of MS/MS are the increased sensitivity (in QQQ, due to reduction of noise) and the increased structural information on the analyte (qTOF) based on the fragmentation pattern (Chernushevich et al., 2001; Ens and Standing, 2005). The detector generates an electrical signal which, after digitisation, is passed on to the data system for evaluation. The signal represents the ion intensity and the time to reach the detector represents a specific  $m/z$  value. The mass spectrometer shows these data as a mass spectrum (Budzikiewicz and Schäfer, 2012).

### 1.3 High-Resolution Mass Spectrometry

Nowadays, low-resolution mass spectrometry (LRMS) is a state-of-the-art technique for the determination of medium to polar organic compounds. The most widely used MS in the quantitative analysis are triple quadrupole mass spectrometers. However, high-resolution mass spectrometers (HRMS) have become increasingly important in recent years by expanding environmental and water monitoring. The analytical window was thus widened and the identification of so far unknown compounds became possible (Heeb et al., 2012; Hollender et al., 2019; Parrilla Vázquez et al., 2018; Richardson and Kimura, 2016; Schmidt, 2018;

Schymanski et al., 2014a), due to the availability of more rugged, sensitive and selective instrumentation (Schymanski et al., 2014). The benefit of HRMS is the detection of full scan mass spectra, which provides accurate mass data while having enough selectivity for environmentally relevant concentrations, even in highly complex environmental samples (Milman and Zhurkovich, 2017; Parrilla Vázquez et al., 2018). Therefore, compounds are measured without previous compound-specific tuning, which enables the screening without using reference standards enables (Brüggen and Schmitz, 2018). However, for an unequivocal identification, the confirmation with reference standards is still required (Purschke et al., 2020).

LC-HRMS has emerged as a powerful detection technique, as the instruments have high mass accuracy ( $\pm 0.001$  Da), high sensitivity and wide mass ranges. That allows the detection of ions formed in the ion source in the selected mass range and the determination of their accurate mass at any point in the chromatogram (Hollender et al., 2017). There are different types of HRMS systems. The HRMS mass analysers most commonly used today are qTOF with a resolution ( $R = \frac{m}{\Delta m}$ , with  $m \triangleq$  mass and  $\Delta m \triangleq$  the difference between two mass peaks) of  $R = 30,000 - 50,000$  (e.g. SCIEX qTOF) and Orbitrap systems with resolutions of  $R = 25,000 - 140,000$  (e.g. Thermo Fisher Orbitrap) depending on the analyser type, the mass range and scan speed (Richardson and Ternes, 2018). Both are typically operated in tandem MS mode with the automated acquisition of fragment ion spectra. The qTOF, which was used in this project (schematically depicted in Figure 1-3), consists of a triple quadrupole and time-of-flight tube.

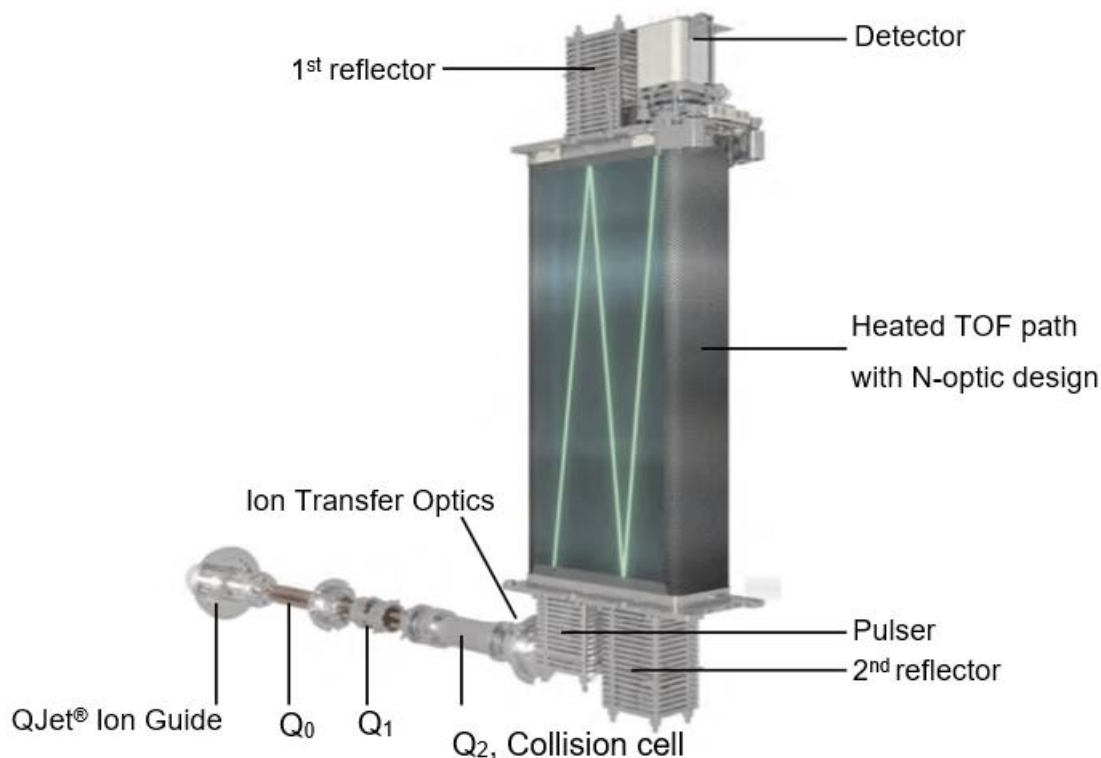


Figure 1-3 – Schematic overview of the x500R qTOF instrument from SCIEX used in this work. © 2017 AB SCIEX

The  $Q_0$ , as well as the ion guide, serve as ion focusing. In the first quadrupole ( $Q_1$ ), the parent ions are chosen. Mass ions are accelerated flying through the four electrodes. By adjusting the voltage, ions with defined  $m/z$  ratio can be selected (De Hoffmann and Stroobant, 2007, pp. 126–142; Hug, 2015, p. 233). If wanted, the parent ions could be fragmented in the second quadrupole ( $Q_2$ , the collision cell) by inert gas (e.g. nitrogen). The use of the collision cell depends on the MS/MS-method (e.g. TOF-MS-scan, IDA, SWATH® or MRM<sup>HR</sup>, see Chapter 3). The time-of-flight path is used to detect the ions by measuring the time they need to transverse the flight path (see Figure 1-3). N-optic design with reflectrons is used for the flight path in this qTOF. The ion beam is reflected by a constant electrostatic field toward the detector. More energetic ions penetrate deeper into the reflection zone and take a longer path to the detector. Less energetic ions penetrate a shorter distance into the reflector and, correspondingly, take a shorter path to the detector. This leads to a reduction of energy differences between the ions and consequently to a higher resolution compared to linear tubes. The separation principle is the speed, respectively, the flying time (Budzikiewicz and Schäfer, 2012). The qTOF combines sensitivity with mass accuracy for both parent and daughter ions. Due to this, the reliable identification of substances, even for low analyte concentrations, is possible (Masiá et al., 2014).

## 1.4 Non-Target Screening

The main outline for LC-HRMS methods consists of three categories: the conventional target analysis, that is used for quantification of target analytes with the help of reference standards, the qualitative suspected-target and non-target screening (NTS). The latter has received much attention in recent years (Brüggen and Schmitz, 2018; Ccancapa-Cartagena et al., 2019; Hollender et al., 2019; Kiefer et al., 2019). Suspected-target screening is performed, having prior structural information of the substance (Krauss et al., 2010; Ruttkies et al., 2019; Samanipour et al., 2017). A list of suspected compounds is necessary, which can be used searching for the exact mass, enabling the advantage of analysing many compounds without the need for a reference standard in a first approximation (Hollender et al., 2017). However, for the identification and confirmation of the structure, a reference standard is still required. The screening method is, therefore, essentially a compromise between target and non-target screening. With NTS, known, previously unrecognised and often unknown substances can be detected (Hug et al., 2014; Schymanski et al., 2014b). NTS approaches do not use any prior information and thus provide a more comprehensive overview of the compounds present in a sample (Peter et al., 2019). Figure 1-4 shows a schematic overview of the data evaluation strategies used in LC-HRMS measurements, including the NTS workflow, to determine all detected ions.

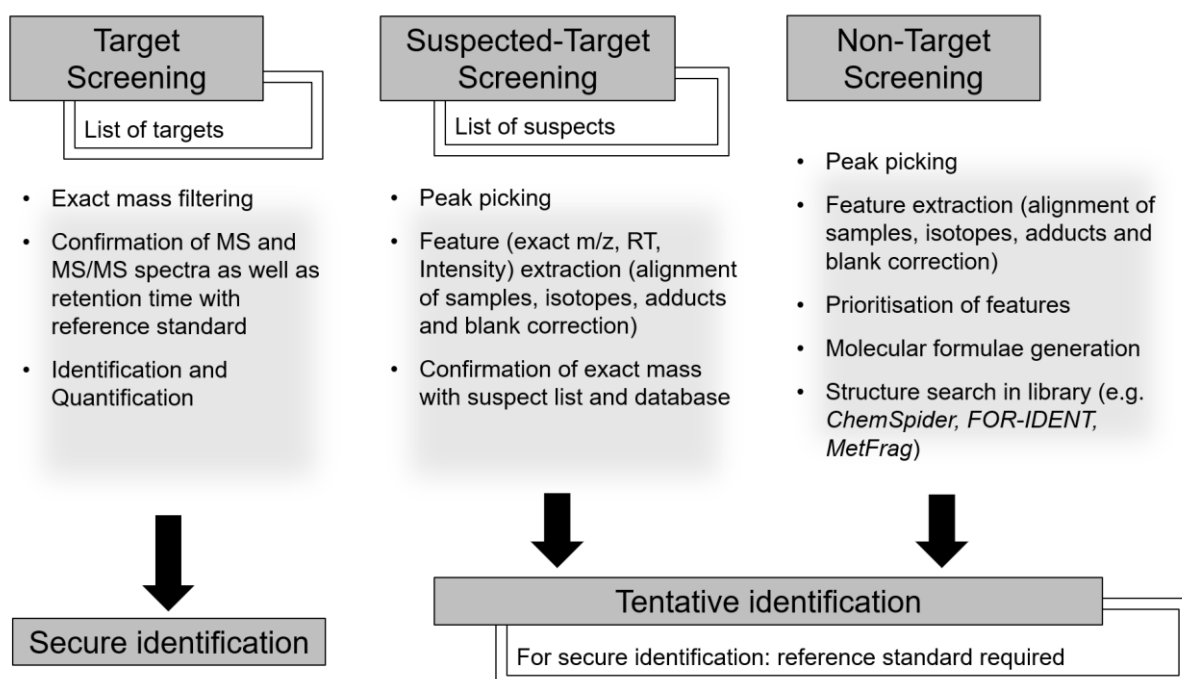


Figure 1-4 – Schematic representation of data evaluation strategies used for LC-HRMS measurements including target screening (reference standards are available), suspected-target screening (prior structural information of the suspects available, but no reference standards are available) and NTS (no previous knowledge and no reference standards are available) based on (Krauss et al., 2010) and (Bletsou et al., 2015).

With peak-picking algorithms, features (three-dimensional data of  $m/z$ , retention time and intensity (Bader et al., 2016)) are extracted, which can be grouped after filtering, blank value subtractions, peak alignments by componentisation (Alygizakis et al., 2019; Bader et al., 2017; Hollender et al., 2017; Köppe et al., 2020; Ruttkies et al., 2019; Schymanski et al., 2014b). However, the large amount of screening data, which are produced with NTS, require a data evaluation concept. Even after filtering or strict thresholds in the parameters of the data processing workflow (e.g. blank subtraction), the number of features is still huge (depending on the matrix). The structure elucidation of all of them is not feasible since it would involve extensive time and effort. Consequently, the selection of features of interest (peak prioritisation) is recommended (Hollender et al., 2017).

Depending on the aims of the study, different prioritisation strategies are applied (Krauss et al., 2019; Samanipour et al., 2017). Many approaches use an intensity-based prioritisation (Aceña et al., 2015; Hohrenk et al., 2020; Ruff et al., 2015; Samanipour et al., 2019; Schymanski et al., 2014b) or a data reduction concerning distinctive isotopic pattern (Gago-Ferrero et al., 2015). Intensity is a crucial parameter to obtain high-quality MS/MS data, which is essential for further identification steps. Furthermore, high intensities are frequently correlated to relevant concentrations. An example of this procedure is presented by Hug et al. (2014), who aimed to find environmentally pertinent substances of wastewater. In that study, high-intensity features and accompanying isotope features were selected for further identification (Hug et al., 2014). Furthermore, statistical procedures for LC-HRMS data are used to characterise, prioritise and identify water contaminants. For example, pattern recognition techniques are used to visualise undiscovered information and patterns like correlations between the in- and outlet of WWTPs to monitor effects in real-time (Itzel et al., 2020). Furthermore, long-time trend analyses based on retrospective data are used to prioritise relevant compounds for further analysis (Creusot et al., 2020). In a study carried out by Schlüsener et al. (2015), LC-HRMS measurements were used as a screening and prioritisation tool on a long-time series of samples from one sampling site that receives municipal and industrial wastewater inputs. This study prioritised features whose intensities varied substantially over time to identify emissions by industrial WWTPs (Schlüsener et al., 2015). Besides, other studies conducted peak prioritisation based on effect-directed analysis (EDA). In EDA, chromatographic fractions associated with specific toxic effects are prioritised for subsequent identification in complex environmental samples (Stütz et al., 2019; Tousova et al., 2018).

In principle, due to high-resolution MS application, tentative identifications are possible by elucidating the molecular formula with the accurate mass and isotope ratios. Fragmentation information of MS/MS spectra help to find structure-based information but sometimes is

insufficient (Purschke et al., 2020). However, the identification of components in NTS requires gathering evidence from many different sources. Usually, a list of candidate formulas is ranked as a function of the accurate mass error obtained by comparing the accurate and the exact mass. The molecular formulae proposed can be searched in commercial databases trying to assign a structure. Some molecular databases, including *ChemSpider* (Royal Society of Chemistry, 2017), *PubChem* (Bolton et al., 2008) and *Metlin* (a metabolite database (Guijas et al., 2018)), also contain MS/MS spectra for data comparison. Beyond spectral libraries, *in silico* fragmentation techniques assist in the structure elucidation of suspected components, e.g. *MetFrag* (Ruttkies et al., 2016). Nevertheless, entirely unknown chemicals, which were by now not reported, cannot be found in such databases. In these cases, further preliminary information utilising complementary techniques to MS, e.g. nuclear magnetic resonance spectroscopy (NMR), as  $^1\text{H-NMR}$  or  $^{13}\text{C-NMR}$ , are needed elucidating the structure (De Vijlder et al., 2018; Wick et al., 2011). Finally, for comparing results from different laboratories, uniform categorisation is required. Schymanski et al. introduced a level system (see Figure 1-5) for the identification of unknown substances analysed by NTS (Schymanski et al., 2014a).

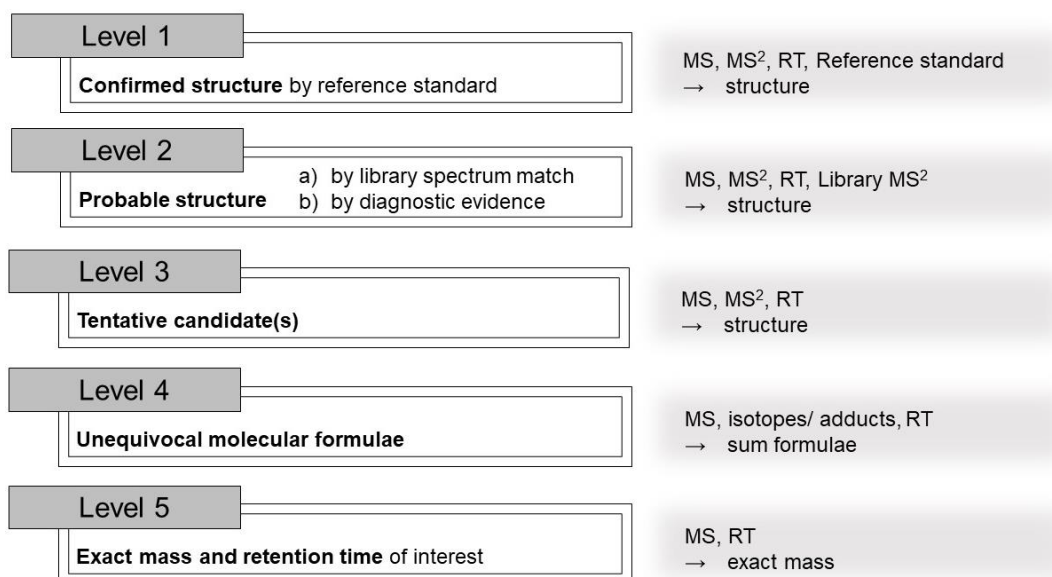


Figure 1-5 – Identification confidence levels in HRMS analysis. Redrafted from Schymanski et al. (Schymanski et al., 2014a).

The probability of identification is increased from level five to one. In level one, for example, MS, MS/MS and retention time information must be confirmed with a reference standard allowing the elucidation of structures (Schymanski et al., 2014; Krauss, Singer and Hollender, 2010). Nevertheless, to identify features, information about the sample being examined is helpful. Such additional information can be properties of substances, occurrences, areas of use, used amounts, possible transformation or by-products in production or application.

## 1.5 Strategies of Non-Target Screening in Industrial Wastewater Analysis

Today, the use of LC-HRMS is the most powerful technique for screening of TrOCs at environmentally relevant concentrations in complex environmental samples. Although most investigations still focus on target screening selectively determining target compounds, there is an increasing number of studies dealing with both, suspect and non-target screening (Brüggen and Schmitz, 2018; Ccancapa-Cartagena et al., 2019; Hollender et al., 2019; Kiefer et al., 2019; Milman and Zhurkovich, 2017). Comparing the TOC of influent samples of industrial WWTPs with the detected number of target analytes, typically 30% remain unknown, which, e.g. arise due to chemical reactions in the wastewater streams. Thus, elucidation by unknown screening is necessary. In literature, applications of NTS for the investigation of wastewater matrices are found for the evaluation of treatment processes and the identification of new, and so far, unknown contaminants. For example, it has been used for the analysis of pharmaceuticals in wastewater influent (Gago-Ferrero et al., 2015), TrOC in wastewater effluent (Hug et al., 2014), assessment of chemical removal at different stages of the wastewater treatment process (Nürenberg et al., 2015). Gomez et al. (2010) and Schollée et al. (2015) identified transformation products of biological wastewater treatment within the WWTP (Gomez et al., 2010; Schollée et al., 2015). Furthermore, Schollée et al. (2018) and Itzel et al. (2020) used NTS to assess wastewater treatment processes by understanding the behaviour of unknown features during ozonation and post-treatment (Itzel et al., 2020; Schollée et al., 2018). Hence, NTS with HRMS has already proven its feasibility in detecting and identifying TrOCs as well as monitoring of wastewater. However, for industrial wastewater analysis using NTS, there are only a few recent studies published in the literature (Iadaresta et al., 2019; Krauss et al., 2019; Ponce-Robles et al., 2018; Purschke et al., 2020). The main reason is the discretion of industrial production processes. Announcements of NTS results may conclude with the elucidation of so far unknown TrOCs and consequently allow to derive information on industrial productions. Furthermore, in comparison to municipal wastewater, industrial wastewater complicates NTS because of more variation due to the production processes. Besides, identification processes may take an unusually long time because substances are in the focus that were not frequently reported and, therefore, not listed in databases.

Nevertheless, the advantages of NTS can also be used here. Prioritisation methods reduce the workload of identification. In addition, the application in the industry offers the advantage of tracing the wastewater flows to locate the production sites that are responsible for the discharge, so that information about unknown structures can also be obtained via internal databases (e.g. element compositions (Purschke et al., 2020)). Furthermore, wastewater



systems can be observed, assessed and subsequently improved without direct identification of TrOCs. Correlations with, for example, online measurement data or long-term trend detections allow conclusions on influences on the production processes even without final structure elucidation of an unknown. However, NTS in industrial wastewaters can also be applied for identification of unknowns as for municipal wastewaters. For example, in recent years, several substances, previously not widely considered to be present in wastewater, were detected via NTS in treated wastewater and the aquatic environment (Hug et al., 2014). All in all, the screening procedure has a significant potential for tasks such as practical evaluation of environmental regulations for assessment of treatment processes, prioritisation of substances for monitoring programmes and evaluation of wastewater quality.

### 1.6 References

- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography - High-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 407, 6289–6299. <https://doi.org/10.1007/s00216-015-8852-6>
- Alygizakis, N.A., Gago-Ferrero, P., Hollender, J., Thomaidis, N.S., 2019. Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater. *J. Hazard. Mater.* 361, 19–29. <https://doi.org/10.1016/j.jhazmat.2018.08.073>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2017. LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes. *Anal. Chem.* 89, 13219–13226. <https://doi.org/10.1021/acs.analchem.7b03037>
- Bader, T., Schulz, W., Lucke, T., 2016. Application of non-target analysis with LC-HRMS for the monitoring of raw and potable water: Strategy and results, in: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 2. ACS Symposium Series, Washington, DC.*, pp. 49–70. <https://doi.org/10.1021/bk-2016-1242.ch003>
- Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. *TrAC - Trends Anal. Chem.* 66, 32–44. <https://doi.org/10.1016/j.trac.2014.11.009>
- Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Sci. Total Environ.* 575, 265–275. <https://doi.org/10.1016/j.scitotenv.2016.09.135>
- Bolton, E.E., Wang, Y., Thiessen, P.A., Bryant, S.H., 2008. Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* 4, 217–241. [https://doi.org/https://doi.org/10.1016/S1574-1400\(08\)00012-1](https://doi.org/https://doi.org/10.1016/S1574-1400(08)00012-1)
- Bristow, A.W.T., 2006. Accurate mass measurement for the determination of elemental formula—A tutorial. *Mass Spectrom. Rev.* 25, 99–111. <https://doi.org/https://doi.org/10.1002/mas.20058>

- Brüggen, S., Schmitz, O.J., 2018. A New Concept for Regulatory Water Monitoring Via High-Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. *J. Anal. Test.* 2, 342–351. <https://doi.org/10.1007/s41664-018-0081-5>
- Budzikiewicz, H., Schäfer, M., 2012. *Massenspektrometrie: Eine Einführung*, 6th ed. Wiley-VCH, Weinheim. <https://doi.org/10.1002/nadc.19990470331>
- Ccanccapa-Cartagena, A., Pico, Y., Ortiz, X., Reiner, E.J., 2019. Suspect, non-target and target screening of emerging pollutants using data independent acquisition: Assessment of a Mediterranean River basin. *Sci. Total Environ.* 687, 355–368. <https://doi.org/10.1016/j.scitotenv.2019.06.057>
- Chernushevich, I. V., Loboda, A. V., Thomson, B.A., 2001. An introduction to quadrupole–time-of-flight mass spectrometry. *J. Mass Spectrom.* 36, 849–865. <https://doi.org/https://doi.org/10.1002/jms.207>
- Creusot, N., Casado-Martinez, C., Chiaia-Hernández, A., Ferrari, B.J.D., Fischer, S., Fu, Q., Munz, N., Singer, H., Stamm, C., Tlili, A., Hollender, J., 2020. Retrospective Screening of High Resolution Mass Spectrometry Archived Digital Samples can Improve Environmental Risk Assessment of Emerging Contaminants: A Case Study on Antifungal Azoles (submitted). *Environ. Int. J.* 139, 105708. <https://doi.org/10.1016/j.envint.2020.105708>
- Crimmins, B.S., Holsen, T.M., 2019. Non-targeted Screening in Environmental Monitoring Programs, in: *Advancements of Mass Spectrometry in Biomedical Research*. pp. 731–741. <https://doi.org/10.1007/978-3-030-15950-4>
- De Hoffmann, E., Stroobant, V., 2007. *Mass Spectrometry-Principles and Applications.*, 3rd ed, Mass spectrometry reviews. John Wiley & Sons, Ltd, West Sussex, England. <https://doi.org/10.1002/mas.20296>
- De Vijlder, T., Valkenburg, D., Lemièrre, F., Romijn, E.P., Laukens, K., Cuyckens, F., 2018. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* 37, 607–629. <https://doi.org/10.1002/mas.21551>
- Dvořák, L., Lederer, T., Jirků, C., J, M., L, N., 2014. Removal of aniline, cyanides and diphenylguanidine from industrial wastewater using a full-scale moving bed biofilm reactor. *Process Biochem.* 49, 102–109. <https://doi.org/https://doi.org/10.1016/j.procbio.2013.10.011>
- Eggen, R.I.L., Hollender, J., Joss, A., Schärer, M., Stamm, C., 2014. Reducing the discharge of micropollutants in the aquatic environment: The benefits of upgrading wastewater treatment plants. *Environ. Sci. Technol.* 48, 7683–7689. <https://doi.org/10.1021/es500907n>
- Ens, W., Standing, K.G., 2005. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol.* 402, 49–78. [https://doi.org/10.1016/S0076-6879\(05\)02002-1](https://doi.org/10.1016/S0076-6879(05)02002-1)
- Fang, N., Yu, S., Ronis, M.J.J., Badger, T.M., 2015. Matrix effects break the LC behavior rule for analytes in LC-MS/MS analysis of biological samples. *Exp. Biol. Med.* 240, 488–497. <https://doi.org/10.1177/1535370214554545>
- Freeling, F., Alygizakis, N.A., von der Ohe, P.C., Slobodnik, J., Oswald, P., Aalizadeh, R., Cirka, L., Thomaidis, N.S., Scheurer, M., 2019. Occurrence and potential environmental risk of surfactants and their transformation products discharged by wastewater treatment plants. *Sci. Total Environ.* 681, 475–487. <https://doi.org/10.1016/j.scitotenv.2019.04.445>
- Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize

- Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* 49, 12333–12341. <https://doi.org/10.1021/acs.est.5b03454>
- Gomez, M.J., Gomez-Ramos, M.M., Malato, O., Mezcua, M., Fernandez-Alba, 2010. Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography- quadrupole-time-of-flight mass spectrometry with an accurate-mas. *J. Chromatogr. A* 1217, 7038–7054. <https://doi.org/http://dx.doi.org/10.1016/j.chroma.2010.08.070>
- Gritter, R.J., Bobbitt, J.M., Schwarting, A.E., 1987. Die Wahl des Phasensystems in der Flüssigkeits-Chromatographie, in: *Einführung in Die Chromatographie*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 87–110. [https://doi.org/10.1007/978-3-642-72789-4\\_3](https://doi.org/10.1007/978-3-642-72789-4_3)
- Guijas, C., Montenegro-Burke, J.R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., Koellensperger, G., Huan, T., Uritboonthai, W., Aisporna, A.E., Wolan, D.W., Spilker, M.E., Benton, H.P., Siuzdak, G., 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* 90, 3156–3164. <https://doi.org/doi:10.1021/acs.analchem.7b04424>.
- Gupta, S. kumar, Bux, F., 2019. Application of Microalgae in Wastewater Treatment Volume 1: Domestic and Industrial Wastewater Treatment, Application of Microalgae in Wastewater Treatment. <https://doi.org/10.1007/978-3-030-13909-4>
- Haag, A.M., Mirzaei, H., Carrasco, M., 2016. Mass Analyzers and Mass Spectrometers, in: *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*. Advances in Experimental Medicine and Biology. Springer, pp. 157–169. [https://doi.org/https://doi.org/10.1007/978-3-319-41448-5\\_7](https://doi.org/https://doi.org/10.1007/978-3-319-41448-5_7)
- Heeb, F., Singer, H., Pernet-Coudrier, B., Qi, W., Liu, H., Longrée, P., Müller, B., Berg, M., 2012. Organic micropollutants in rivers downstream of the megacity Beijing: Sources and mass fluxes in a large-scale wastewater irrigation system. *Environ. Sci. Technol.* 46, 8680–8688. <https://doi.org/10.1021/es301912q>
- Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.* 92, 1898–1907. <https://doi.org/https://doi.org/10.1021/acs.analchem.9b04095>
- Holčapek, M., Jirásko, R., Lísa, M., 2010. Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules. *J. Chromatogr. A* 1217, 3908–3921. <https://doi.org/10.1016/j.chroma.2010.02.049>
- Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* 51, 11505–11512. <https://doi.org/10.1021/acs.est.7b02184>
- Hollender, J., van Bavel, B., Dulio, V., Farmen, E., Furtmann, K., Koschorreck, J., Kunkel, U., Krauss, M., Munthe, J., Schlabach, M., Slobodnik, J., Stroomberg, G., Ternes, T., Thomaidis, N.S., Togola, A., Tornero, V., 2019. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environ. Sci. Eur.* 31. <https://doi.org/10.1186/s12302-019-0225-x>
- Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* 184, 25–32. <https://doi.org/10.1016/j.envpol.2013.07.048>
- Hug, H., 2015. *Instrumentelle Analytik Theorie und Praxis*, 3rd ed. Verlag Europa-Lehrmittel, Haan-Gruiten.

- Iadaresta, F., Carlsson, J., Eklund, J., Avagyan, R., Östman, C., 2019. Strategies Towards Suspect and Non-target Screening of Chemicals in Clothing Textiles by Reverse Phase Liquid Chromatography–hybrid Linear Ion Trap Orbitrap Mass Spectrometry.
- Itzel, F., Baetz, N., Hohrenk, L.L., Gehrmann, L., Antakyali, D., Schmidt, T.C., Tuerk, J., 2020. Evaluation of a biological post-treatment after full-scale ozonation at a municipal wastewater treatment plant. *Water Res.* 170, 115316. <https://doi.org/10.1016/j.watres.2019.115316>
- Keller, B.O., Sui, J., Young, A.B., Whittall, R.M., 2008. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* 627, 71–81. <https://doi.org/10.1016/j.aca.2008.04.043>
- Kessler, R.W., 2006. *Prozessanalytik Strategien und Fallbeispiele aus der industriellen Praxis*, 1st ed. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. <https://doi.org/978-3-527-31196-5>
- Kiefer, K., Müller, A., Singer, H., Hollender, J., 2019. New relevant pesticide transformation products in groundwater detected using target and suspect screening for agricultural and urban micropollutants with LC-HRMS. *Water Res.* 165, 114972. <https://doi.org/10.1016/j.watres.2019.114972>
- Köppe, T., Jewell, K.S., Dietrich, C., Wick, A., Ternes, T.A., 2020. Application of a non-target workflow for the identification of specific contaminants using the example of the Nidda river basin. *Water Res.* 178. <https://doi.org/10.1016/j.watres.2020.115703>
- Krauss, M., Hug, C., Bloch, R., Schulze, T., Brack, W., 2019. Prioritising site-specific micropollutants in surface water from LC-HRMS non-target screening data using a rarity score. *Environ. Sci. Eur.* 31, 45. <https://doi.org/10.1186/s12302-019-0231-z>
- Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: From target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–951. <https://doi.org/10.1007/s00216-010-3608-9>
- Kumari, V., Tripathi, A.K., 2019. Characterization of pharmaceuticals industrial effluent using GC–MS and FT-IR analyses and defining its toxicity. *Appl. Water Sci.* 9, 1–8. <https://doi.org/10.1007/s13201-019-1064-z>
- Leendert, V., Van Langenhove, H., Demeestere, K., 2015. Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments. *TrAC - Trends Anal. Chem.* 67, 192–208. <https://doi.org/10.1016/j.trac.2015.01.010>
- Masiá, A., Campo, J., Blasco, C., Picó, Y., 2014. Ultra-high performance liquid chromatography–quadrupole time-of-flight mass spectrometry to identify contaminants in water: An insight on environmental forensics. *J. Chromatogr. A* 1345, 86–97. <https://doi.org/10.1016/j.chroma.2014.04.017>
- Matuszewski, B.K., Constanzer, M.L., Chavez-Eng, C.M., 2003. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC–MS/MS. *Anal. Chem.* 75, 3019–3030. <https://doi.org/10.1021/ac020361s>
- Mellon, F.A., Bennett, R.N., Holst, B., Williamson, G., 2002. Intact glucosinolate analysis in plant extracts by programmed cone voltage electrospray LC/MS: Performance and comparison with LC/MS/MS methods. *Anal. Biochem.* 306, 83–91. <https://doi.org/10.1006/abio.2002.5677>
- Meyer, V.R., 2009. *Praxis der Hochleistungs-Flüssigchromatographie*, 10th ed. Wiley-VCH, Weinheim.
- Milman, B.L., Zhurkovich, I.K., 2017. The chemical space for non-target analysis. *Trends Anal. Chem.* 97, 179–187. <https://doi.org/10.1016/j.trac.2017.09.013>

- Murray, K.K., Boyd, R.K., Eberlin, M.N., Langley, G.J., Li, L., Naito, Y., 2013. Standard definitions of terms relating to mass spectrometry. *J. Am. Soc. Mass Spectrom.* 85, 1515–1609. <https://doi.org/http://dx.doi.org/10.1351/PAC-REC-06-04-06>
- Nagy, K., Vékey, K., 2008. Chapter 5 - Separation methods, in: Vékey, Károly, Telekes, A., Vertes, A. (Eds.), *Medical Applications of Mass Spectrometry*. Elsevier, Amsterdam, pp. 61–92. <https://doi.org/https://doi.org/10.1016/B978-044451980-1.50007-0>
- Neue, U.D., Phoebe, C.H., Tran, K., Cheng, Y.F., Lu, Z., 2001. Dependence of reversed-phase retention of ionizable analytes on pH, concentration of organic solvent and silanol activity. *J. Chromatogr. A* 925, 49–67. [https://doi.org/10.1016/S0021-9673\(01\)01009-3](https://doi.org/10.1016/S0021-9673(01)01009-3)
- Nürenberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography - high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>
- O’Gara, J.E., Walsh, D.P., Alden, B.A., Casellini, P., Walter, T.H., 1999. Systematic study of chromatographic behavior vs alkyl chain length for HPLC bonded phases containing an embedded carbamate group. *Anal. Chem.* 71, 2992–2997. <https://doi.org/10.1021/ac9900331>
- Oyetibo, G.O., Chien, M.-F., Ikeda-Ohtsubo, W., Suzuki, H., Obayori, O.S., Adebusoye, S.A., Ilori, M.O., Amund, O.O., Endo, G., 2017. Biodegradation of crude oil and phenanthrene by heavy metal resistant *Bacillus subtilis* isolated from a multi-polluted industrial wastewater creek. *Int. Biodeterior. Biodegradation* 120, 143–151. <https://doi.org/https://doi.org/10.1016/j.ibiod.2017.02.021>
- Özkaraoğlu, E.B., Akbal, F., Kuleyin, A., 2018. Potential reuse of treated industrial wastewater in agriculture: Textile wastewater, in: *International Scientific Journal Mechanisation in Agriculture*. pp. 138–140.
- Parrilla Vázquez, P., Lozano, A., Ferrer, C., Martínez Bueno, M.J., Fernández-Alba, A.R., 2018. Improvements in identification and quantitation of pesticide residues in food by LC-QTOF using sequential mass window acquisition (SWATH®). *Anal. Methods* 10, 2821–2833. <https://doi.org/10.1039/c8ay00678d>
- Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Non-Target High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. *Environ. Sci. Technol.* 53, 12257–12268. <https://doi.org/10.1021/acs.est.9b04481>
- Ponce-Robles, L., Oller, I., Agüera, A., Trinidad-Lozano, M.J., Yuste, F.J., Malato, S., Perez-Estrada, L.A., 2018. Application of a multivariate analysis method for non-target screening detection of persistent transformation products during the cork boiling wastewater treatment. *Sci. Total Environ.* 633, 508–517. <https://doi.org/10.1016/j.scitotenv.2018.03.179>
- Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Sci. Total Environ.* 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>
- Rager, J.E., Strynar, M.J., Liang, S., McMahan, R.L., Richard, A.M., Grulke, C.M., Wambaugh, J.F., Isaacs, K.K., Judson, R., Williams, A.J., Sobus, J.R., 2016. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. *Environ. Int.* 88, 269–280. <https://doi.org/10.1016/j.envint.2015.12.008>
- Richardson, S.D., Kimura, S.Y., 2016. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* 88, 546–582. <https://doi.org/10.1021/acs.analchem.5b04493>

- Richardson, S.D., Ternes, T.A., 2018. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* 90, 398–428. <https://doi.org/10.1021/acs.analchem.7b04577>
- Robles-Molina, J., Lara-Ortega, F.J., Gilbert-López, B., García-Reyes, J.F., Molina-Díaz, A., 2014. Multi-residue method for the determination of over 400 priority and emerging pollutants in water and wastewater by solid-phase extraction and liquid chromatography-time-of-flight mass spectrometry. *J. Chromatogr. A* 1350, 30–43. <https://doi.org/10.1016/j.chroma.2014.05.003>
- Rohner, T.C., Lion, N., Girault, H.H., 2004. Electrochemical and theoretical aspects of electrospray ionisation. *Phys. Chem. Chem. Phys.* 6, 3056–3068. <https://doi.org/10.1039/b316836k>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Rüdel, H., Körner, W., Letzel, T., Neumann, M., Nödler, K., Reemtsma, T., 2020. Persistent, mobile and toxic substances in the environment: a spotlight on current research and regulatory activities. *Environ. Sci. Eur.* 32. <https://doi.org/DOI:10.1186/s12302-019-0286-x>.
- Ruff, M., Mueller, M.S., Loos, M., Singer, H.P., 2015. Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry - Identification of unknown sources and compounds. *Water Res.* 87, 145–154. <https://doi.org/10.1016/j.watres.2015.09.017>
- Ruttkies, C., Schymanski, E.L., Williams, A.J., Krauss, M., 2019. Supporting non-target identification by adding hydrogen deuterium exchange MS / MS capabilities to MetFrag. *Anal. Bioanal. Chem.* 411, 4683–4700. <https://doi.org/https://doi.org/10.1007/s00216-019-01885-0>
- Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 1–16. <https://doi.org/10.1186/s13321-016-0115-9>
- Samanipour, S., Kaserzon, S., Vijayasathy, S., Jiang, H., Choi, P., Reid, M.J., Mueller, J.F., Thomas, K. V., 2019. 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. <https://doi.org/10.1016/j.talanta.2018.11.039>
- Samanipour, S., Reid, M.J., Thomas, K. V., 2017. Statistical Variable Selection: An Alternative Prioritization Strategy during the Nontarget Analysis of LC-HR-MS Data. *Anal. Chem.* 89, 5585–5591. <https://doi.org/10.1021/acs.analchem.7b00743>
- Schlüsener, M.P., Kunkel, U., Ternes, T.A., 2015. Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants. *Environ. Sci. Technol.* 49, 14282–14291. <https://doi.org/10.1021/acs.est.5b03926>
- Schmidt, T.C., 2018. Recent trends in water analysis triggering future monitoring of organic micropollutants. *Anal. Bioanal. Chem.* 410, 3933–3941. <https://doi.org/10.1007/s00216-018-1015-9>
- Schollée, J.E., Bourgin, M., von Gunten, U., Mc Ardell, C.S., Hollender, J., 2018. Non-target screening to trace ozonation transformation products in a wastewater treatment train including different post-treatments. *Water Res.* 142, 267–278. <https://doi.org/10.1016/j.watres.2018.05.045>
- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* 87, 12121–12129. <https://doi.org/10.1021/acs.analchem.5b02905>

- Schollée, J.E., Schymanski, E.L., Hollender, J., 2016. Statistical Approaches for LC-  
HRMS Data to Characterize, Prioritize, and Identify Transformation Products from  
Water Treatment Processes. *ACS Symp. Ser.* 1241, 45–65.  
<https://doi.org/10.1021/bk-2016-1241.ch004>
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J.,  
2014a. Identifying small molecules via high resolution mass spectrometry:  
Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098.  
<https://doi.org/10.1021/es5002105>
- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés  
Vidal, C., Hollender, J., 2014b. Strategies to characterize polar organic contamination  
in wastewater: Exploring the capability of high resolution mass spectrometry. *Environ.  
Sci. Technol.* 48, 1811–1818. <https://doi.org/10.1021/es4044374>
- Stütz, L., Leitner, P., Schulz, W., Winzbacher, R., 2019. Identification of genotoxic  
transformation products by effect-directed analysis with high-performance thin-layer  
chromatography and non-target screening. *J. Planar Chromatogr. - Mod. TLC* 32.  
<https://doi.org/https://doi.org/10.1556/1006.2019.32.3.1>
- Thelen, T., Rösen, W., 2017. Reversed Phase-Chromatographie (RP-HPLC) [WWW  
Document]. URL <http://hplc-saeule.de/rp-hplc-mit-gebundenen-phasen/> (accessed  
6.14.17).
- Tousova, Z., Froment, J., Oswald, P., Slobodník, J., Hilscherova, K., Thomas, K. V,  
Tollefsen, K.E., Reid, M., Langford, K., Blaha, L., 2018. Identification of algal growth  
inhibitors in treated waste water using effect-directed analysis based on non-target  
screening techniques. *J. Hazard. Mater.* 258, 494–502.  
<https://doi.org/10.1016/j.jhazmat.2018.05.031>
- Trufelli, H., Palma, P., Famiglioni, G., Cappiello, A., 2011. An overview of matrix effects in  
liquid chromatography–mass spectrometry. *Mass Spectrom. Rev.* 30, 491–509.  
<https://doi.org/10.1002/mas.20298>
- Van der Heeft, E., Bolck, Y.J.C., Beumer, B., Nijrolder, A.W.J.M., Stolker, A.A.M., Nielen,  
M.W.F., 2009. Full-Scan Accurate Mass Selectivity of Ultra-Performance Liquid  
Chromatography Combined with Time-of-Flight and Orbitrap Mass Spectrometry in  
Hormone and Veterinary Drug Residue Analysis. *J. Am. Soc. Mass Spectrom.* 20,  
451–463. <https://doi.org/10.1016/j.jasms.2008.11.002>
- Wick, A., Wagner, M., Ternes, T.A., 2011. Elucidation of the Transformation Pathway of  
the Opium Alkaloid Codeine in Biological Wastewater Treatment. *Environ. Sci.  
Technol.* 45, 33374–3385. <https://doi.org/https://doi.org/10.1021/es103489x>
- Wortberg, M., Kurz, J., 2019. Analytics 4.0: Online wastewater monitoring by GC and  
HPLC. *Anal. Bioanal. Chem.* 411, 6783–6790. <https://doi.org/10.1007/s00216-019-02065-w>

## **Chapter 2 Scope and Aims of the Thesis**

From the wastewater management's perspective, the comprehensive monitoring of industrial wastewater is necessary as trace organic compounds (TrOCs) might affect the quality of wastewater treatment and the environment. Untargeted approaches allow the monitoring of unknown or unexpected compounds and thus, early detection of possible risks for the water treatment. Furthermore, the fate and the behaviour of TrOCs during wastewater treatment processes (e.g. biological sewage treatment) are of high interest to assess the performance of these processes based on all detectable information. The monitoring of wastewater streams or the optimisation of different operating conditions in WWTPs is the required field of application. In non-target screening (NTS), several thousands of features (three-dimensional data of exact mass-to-charge ratio, retention time and intensity) are usually detectable within a single sample. This makes manual reviewing no longer a reasonable option. Instead, automated algorithms are needed to prioritise relevant features from the wealth of data by performing data reduction.

This thesis aims at the development of prioritisation strategies for NTS using liquid chromatography coupled to high-resolution time-of-flight mass spectrometry (LC-HR-qTOF-MS). The additional focus of this work is on industrial wastewater for identifying TrOCs and enabling closer monitoring of sewage. Moreover, the investigation of a more effective evaluation of treatment procedures is part of the presented work. The following scheme (see Figure 2-1) visualises the overall topic of the thesis and highlights the contribution of the individual chapters in the thesis context.



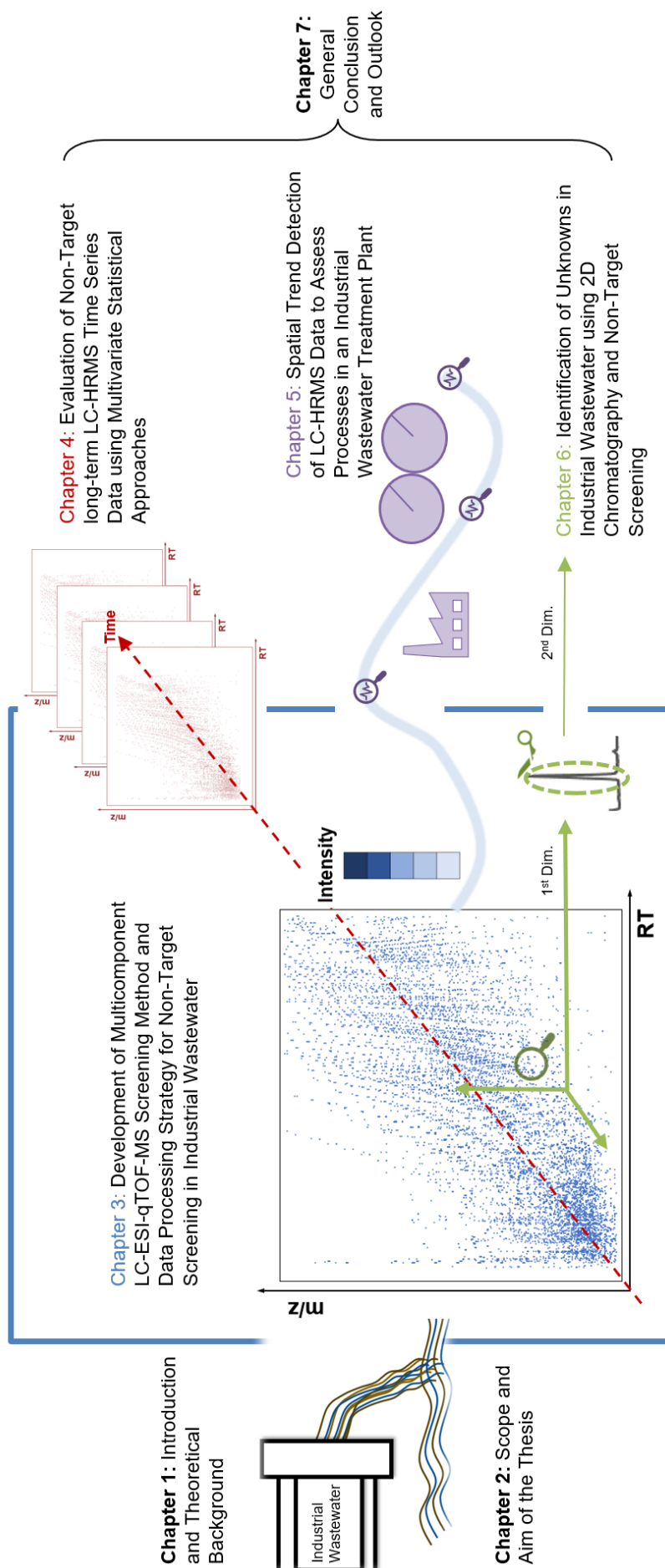


Figure 2-1 – Graphical overview of the content of this thesis.

**Chapter 3** deals with the development of the LC-ESI-qTOF-MS method for NTS in industrial wastewater. According to Figure 2-1, the developed method provides the basis of this thesis. The focus is on the development of a multicomponent LC-HRMS analysis method, which covers a wide polarity range of compounds in industrial wastewater. Furthermore, in NTS, several thousands of features are detected, making data processing algorithms necessary to process the wealth of data. Therefore, in this chapter, the processing method of NTS is developed and validated for authentic wastewater samples.

The following chapters concern prioritisation methods for data reduction, taking relevant TrOCs in industrial wastewater into focus. Screening results of high-resolution mass spectrometry (HRMS) cover a significant amount of data. Thousands of features are detected in one full scan run. Therefore, peak prioritisation plays a crucial role in the NTS of complex samples to focus the elucidation. In **chapter 4**, the compounds are prioritised by their temporal trend. Features, following a predefined relevant temporal pattern, are extracted for further identification procedures. Thus, elucidation efforts are focused on potentially relevant substances. The present work shows the development of a computational workflow, which is capable of detecting compounds that increase or decrease in intensity over time. The developed approach is based on multivariate and univariate statistical tests. While this study limited the determination regarding time trends, a spatial correlation to time series is assigned in **chapter 5**. Data processing strategies are developed for assessing water treatment processes using the information of all detectable compounds. Features of wastewater treatment plant (WWTP) influent are compared with those of the WWTP effluent over five months (November to March) to assess the treatment process of the WWTP over time.

Another prioritisation method is introduced in **chapter 6**. The section contains a comparative study regarding the structure elucidation of so-called 'known unknowns' in routine wastewater analysis using liquid chromatography coupled to ultra-violet detection (LC-UV). The target analysis of industrial wastewater samples using LC-UV delivers insufficient data for structure identification. Therefore, it is demonstrated how these unknowns can be identified using offline two-dimensional chromatography coupled to HRMS with NTS (2D LC-UV x LC-HRMS). Peaks of interest are manually fractionated in the first dimension (LC-UV) followed by the analysis of the eluate fraction in the second dimension (HRMS) with non-target data evaluation.

In **chapter 7**, the general conclusions of this work and an outlook for investigations in further studies arising from the present thesis are described.

## **Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater**

### **3.1 Introduction**

Wastewater is commonly treated in wastewater treatment plants (WWTP) to reduce the contamination load into the environment. Due to complex matrices, transformation reactions and concentration reductions, a selective, highly sensitive analytical method is required for the analysis of trace organic compounds (TrOC) in industrial wastewater. Liquid chromatography (LC) coupled to triple quadrupole mass spectrometry with electrospray ionisation (LC-ESI-MS/MS) is the method of choice for the determination of TrOC (Appa et al., 2018; Marta-Sanchez et al., 2018; Pérez-Fernández et al., 2017; Pozo et al., 2006). The high selectivity by multi reaction monitoring (MRM), as well as excellent sensitivity, enables trace environmental analysis. However, LC-MS/MS used in MRM mode is related to target analysis and requires reference standards during method development. Therefore, screening of unknowns is impossible, which causes a restriction regarding wastewater monitoring.

Recent developments in high-resolution mass spectrometry (HRMS) coupled to LC enabled new possibilities in wastewater monitoring, especially in the analysis of unexpected and unknown components without having any prior information (Hollender et al., 2017; Krauss et al., 2010). Non-target screening (NTS) methods offer the opportunity to monitor wastewater streams, to assess water treatment processes and to identify contaminants at environmentally relevant concentrations. In LC-HRMS, the samples are analysed using full scan methods to capture all present contaminants of the sample separable and ionisable with the current method. More reliable identification by accurate MS is obtained when the full scan is combined with MS/MS enabling fragmentation. The combination is achieved by, e.g. quadrupole/time-of-flight MS (qTOF-MS), in which fragments are produced in collision cell and measured in time-of-flight tube. Standard LC-HR-MS/MS fragmentation modes are data-dependent (DDA) and data-independent acquisition (DIA). DDA is the more commonly applied method, whereas DIA MS techniques developed over the past few years (Wang et al., 2019; Zhu et al., 2014). Both DDA and DIA first acquire a full scan followed by one or multiple MS/MS acquisition scans. DDA typically generates MS/MS data for only the most abundant mass peaks observed in an LC-MS analysis. As a consequence, DDA might result in loss of MS/MS spectra for relevant peaks with lower intensity, e.g. due to ion suppression by matrix

effects (Richardson and Ternes, 2018). In comparison, in DIA, all precursor ions of full scan mode are selected to generate MS/MS spectra. However, the direct link between fragments and precursor ions is more complicated and sometimes even impossible (Brüggen and Schmitz, 2018). New techniques, as variable data-independent acquisitions (e.g. Sequential Window Acquisition of all Theoretical fragment-ion spectra, SWATH®) isolate compounds in selected isolation width windows (e.g. 21 Da) to reduce the number of potentially interfering compounds and enabling the link back to full scan MS (Bonner and Hopfgartner, 2019; Wang et al., 2019).

Regarding this, unknown screening using LC-HRMS has become an analytical method of increasing importance for water monitoring of TrOC (Hollender et al., 2017; Schmidt, 2018). As LC-HR-MS/MS screening methods produce large amounts of raw-data, data treatment concepts are necessary to extract meaningful and valuable information. Recently, many data handling ideas for NTS have emerged (Fiorino et al., 2019; Geppert et al., 2010; Hohrenk et al., 2020; Horlacher et al., 2016; Park et al., 2018; Saurina and Sentellas, 2019). The data handling workflow can be divided into two steps: data processing and data analysis (Katajamaa and Orešič, 2007). During the data processing step, raw data are treated by signal processing methods. Filter procedures as background subtraction, feature extraction, feature alignment and feature normalisation are part of the data processing (Bader et al., 2017; Hohrenk et al., 2020; Köppe et al., 2020; Schulz et al., 2014).

A further challenge represents the analysis of more complex samples like wastewaters, as they can contain high concentrations of matrix interferences. These interferences need to be removed or reduced before injection on the LC-HRMS system. Therefore, the development of sample preparation is the very first challenge of establishing a new analytical method. The preparation technique should be able to keep the full range of compounds with different chemical properties but preventing the system against particle entry which could interfere with the analysis. In qualitative screening methods, there are no required detection limits to be reached. Nevertheless, enough sensitivity is essential. However, general precautionary values are published, which until now, are not directly applicable to wastewater (Hinnenkamp et al., 2019). For NTS, only generic sample preparations are applied, which allows a comprehensive view of any potential compounds of interest but can lead to the loss of compounds (Nürnberg et al., 2015). Besides, sample preparation should be related to the studied matrix. For example, the matrix of the influent and the effluent of a WWTP varies significantly having a different influence on ion intensity suppression and enhancement in mass spectrometry (Nürnberg et al., 2015). Consequently, these matrix effects can lead to a misleading interpretation of intensity variances but partially be reduced during sample preparation.

Therefore, the objective of the present work is the development of an NTS method with reliable data processing using LC-HRMS (qTOF) to expand the analysis and monitoring of industrial wastewater. The screening method expands water monitoring by analysing targets, environmentally suspected contaminants and unknown compounds. For validation, wastewater samples were analysed, applying the developed non-target workflow.

## 3.2 Experimental Section

### 3.2.1 Chemicals and Reagents

Isotope-labelled (deuterated) compounds (ISTD) and target analytes used in this work are alphabetically listed with chemical abstract service (CAS) numbers in the Appendix (see Table-A 3-1). Standard solutions were prepared in methanol and stored at 4–8 °C. For measurement, they were diluted in ultra-pure water containing 0.1% formic acid. ISTD and targets were all of the quality grades suitable for trace analysis, purity  $\geq$  95%. Methanol and acetonitrile with LC-MS grade from Honeywell Riedel-de-Haën™ (Seelze, Germany) were utilised. For ultra-pure water, a Milli-Q® (Q-PoD®) water system from Merck KGaA (Darmstadt, Germany) was used. Formic acid 99% was purchased from Fisher Chemical (Geel, Belgium) and ammonium formate 10 M from Sigma-Aldrich (Steinheim, Germany).

### 3.2.2 Sample Materials

24-h composite flow industrial wastewater samples were used covering five relevant matrices of the industrial WWTP: the WWTP influent (see Figure 3-1, SP 1 to 3), the WWTP effluent (cf. Figure 3-1, SP 5) and one matrix within the WWTP (see Figure 3-1, SP 4). For method development and data evaluation validation, 54 samples, comprising the five mentioned wastewater samples (see Figure 3-1) and ultra-pure water, were spiked with 70 target analytes (cf. Table-A 3-1) at several concentration levels (see Table-A 3-2). The spiked substances were selected to cover the measured mass and retention time range. The developed method was applied to the monitoring of 100 un-spiked samples over two months to prove the suitability of the method. Previously, the samples were declared positive for at least one of the targets of Table-A 3-1 in the Appendix (using routine analysis of LC-UV).

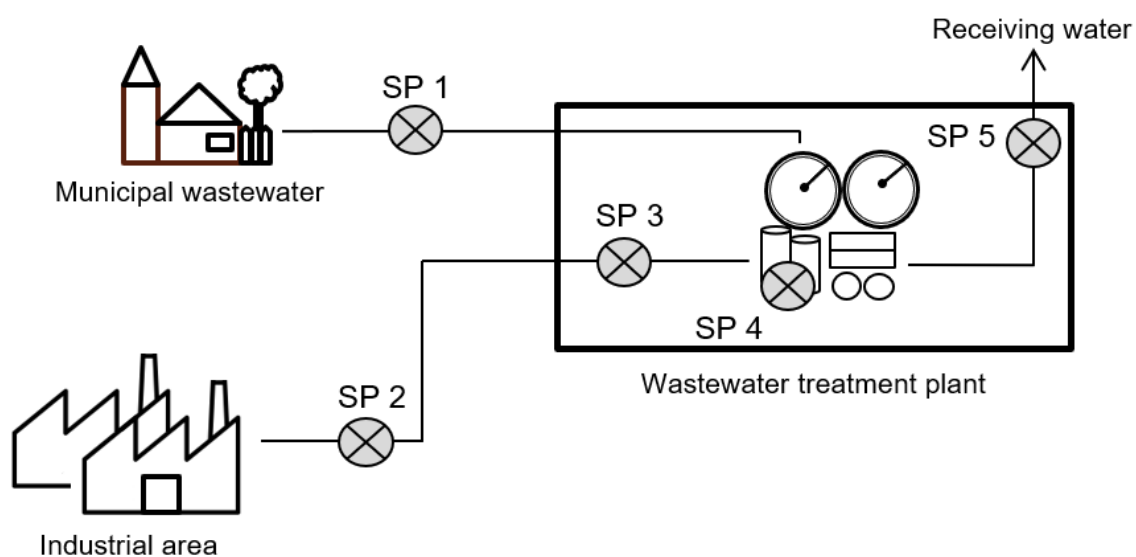


Figure 3-1 – Schematic representation of sampling sites (SP) of used samples.

### 3.2.3 Quality Control

A quality control (QC) was prepared in Eluent A of LC-HRMS (see 3.2.5) using 70 target analytes ( $100 \mu\text{g L}^{-1}$ , see Table-A 3-1). The QC standard was analysed with both ionisation modes at the beginning and the end of each sequence. The analytes of the QC sample should agree with full identification requirements: (1) at least two ions (e.g. precursor and fragment ions), (2) a mass error of  $\leq 5$  ppm for at least one of the ions, (3) an ion ratio within  $\pm 30\%$  and (4) a retention time within  $\pm 0.1$  min tolerance window. Solvent blanks (eluent A, 1:1, V/V) were used to overcome carry-over by measuring them after each fifth sample within the sequence. Furthermore, the LC-HRMS stability across the sequence was evaluated by reviewing the spiked ISTD (exemplarily shown in Appendix section 3.6.1.3). Negative effects such as loss of sensitivity or increase in mass error over time would impair the feature classification and thus can result in poor repeatability.

### 3.2.4 Sample Preparation

Sample preparation took place three days after spiking the wastewater samples. Centrifugation for one minute at 2500 rpm (Centrifuge, 400 Function Line, Heraeus, Hanau, Germany) and dilution using mobile phase A (see 3.2.5) were tested to reduce the introduction of the suspended matter of the (un-)spiked samples (cf. 3.2.2) into the LC system. Before analysis, the (un-)spiked samples, blanks and QC were spiked with an internal standard mixture of 13 ISTD (see Table-A 3-2).

### 3.2.5 Sample Storage

The stability of spiked wastewater samples was tested during different storage conditions during method development. Aliquots (each 1 mL) of the spiked samples were stored for one week at 8 °C and for one week at -22 °C. Before analysis, samples were thawed for 30 minutes. The long-term stability was checked by storing the samples over 6 and 12 months in the freezer at -22 °C.

### 3.2.6 LC-ESI-qTOF-MS Conditions

Samples were analysed in positive and negative ESI. The chromatographic separation was performed using an Agilent 1290 Infinity HPLC system (consisting of vacuum degasser, column temperature control, autosampler and binary pump; Agilent Technologies, Waldbronn, Germany). For stationary phase, four modified LC C18 columns were tested, which could be suitable for the separation of analytes in the difficult wastewater samples. A detailed overview of the used columns is presented in Table 3-1. The corresponding mobile phase, varying in formic acid and ammonium formate concentration as well as in applied solvent (methanol or acetonitrile), were tested for each column. All applied mobile phases are listed in the Appendix (see Table-A 3-8).

Table 3-1 – List of tested columns with their specification and properties (TMS  $\triangleq$  the functional group trimethylsilyl).

<b>Column</b>	<b>Kinetex</b>	<b>Synergi</b>	<b>Raptor Biphenyl</b>	<b>Raptor</b>	<b>Ultra Aqueous</b>
<b>Specification</b>	C18 with TMS end-capping	C18 with ether-linked phenyl and polar end-capping	C18 with biphenyl ligands	Proprietary polar modified, functionally bonded C18	
<b>Column internal diameter in mm</b>	2.1	2	2.1	2.1	
<b>Column length in mm</b>	100	100	100	100	
<b>Particle size in <math>\mu\text{m}</math></b>	2.6	2.5	2.7	3	
<b>Pore size in <math>\text{Å}</math></b>	100	100	90	100	
<b>Distributor</b>	Phenomenex Inc. (Aschaffenburg, Germany)		Restek GmbH (Bad Homburg v. d. Höhe, Germany)		

For the development of data processing strategy and MS fragmentation method, the HPLC was equipped with the polar modified and functionally bonded Raptor™ Ultra Aqueous C18 analytical column (100 mm x 2.1 mm, 3.0  $\mu\text{m}$  particle size) with Restek™ Trident cartridge (10 x 2.1 mm) and a filter (2 mm, 0.5  $\mu\text{m}$ ; Restek GmbH, Bad Homburg v. d. Höhe, Germany). The (pre-)column was changed depending on the peak shape of the analytes of the quality

control (see 3.2.3). Gradient elution was performed with Milli-Q® water as mobile phase A and methanol as mobile phase B, both spiked with 0.1% formic acid. Solvent gradient elution was used. In the optimised analysis method, the initial mobile phase composition (0% B) was held constant for 0.5 min, followed by an increase in B to 10% within 0.5 min, then to 90% in 19 min, kept there for 6.0 min and finally a decrease to initial mobile phase composition (0% B), which was kept for 6.0 min. The column oven was adjusted to 55 °C. The flow rate and the injection mode were set to 500  $\mu\text{L min}^{-1}$  and 5  $\mu\text{L}$ .

The LC system was connected to hybrid quadrupole time-of-flight mass spectrometer x500R qTOF (SCIEX GmbH, Darmstadt, Germany). For ionisation, the electrospray ion source Turbo V™ (SCIEX GmbH, Darmstadt, Germany) was used. Full scan HRMS data were recorded within a mass-to-charge ratio ( $m/z$ ) ranging from 70 to 950 for each sample. For LC-HR-MS/MS analysis, both data-dependent acquisition (DDA, IDA Information Dependent Acquisition from SCIEX) and variable data-independent acquisitions (vDIA, SWATH®, using 'variable window tool' from SCIEX, detailed description of vDIA is found in 0) were tested. For the data processing strategy development, DDA was used containing a TOF-MS survey of 100 ms and up to ten intensity-based data-dependent TOF-MS/MS scans in each cycle (0.56 s) covering a mass range of  $m/z$  30 – 700. Precursor ions exceeding an intensity threshold of 50 counts per second (cps) were the essential requirement for DDA, determining the ten most intense ions. Ions, as well as their isotopes, were excluded from DDA for a period of 4 s after three occurrences. The fragmentation was performed with collision energy (CE) of 35 V and a collision energy spread (CES) of 15 V. Accurate mass measurements were obtained by using an automated recalibration delivery system, which provides mass correction (CDS, ESI positive calibration solution, 5049910 and ESI negative calibration solution, 5042913, SCIEX GmbH, Darmstadt, Germany). The LC flow was directed to waste from 0.1 to 0.9 min enabling MS calibration and from 25.0 to 31.0 min protecting the HRMS system by a post-column diversion valve (Rheodyne, Darmstadt, Deutschland). Experimental details for target compounds used for screening method development are found in the Appendix (see Table-A 3-1).

### 3.2.7 Data Processing and Data Analysis

The full scan data were processed using the SCIEX OS software '*Non-targeted Screening*' workflow (Version 1.4.1, SCIEX GmbH, Darmstadt; Germany; parameters are listed in Appendix Table-A 3-3). The NTS processing algorithm used in this workflow is not published. It consists of a customisable workflow, including criteria for peak picking, blank correction, formulae prediction and an automated library and database search for identification purposes.



For formulae identification, the Formula Finder of *SCIEX OS* Software (SCIEX GmbH, Darmstadt, Germany) was used. For an elemental composition assignment carbon, hydrogen, nitrogen, oxygen, chlorine, fluorine, phosphorus and sulphur atoms (based on production plants producing the industrial wastewater) with a mass tolerance of < 5 ppm were considered based on the typical element composition of the chemicals produced in the production plants. Furthermore, the library searches based on proposed formulae were performed according to the parameters of Table-A 3-3 d) and e) in the Appendix. Additionally, online databases, e.g. *ChemSpider* and *FOR-IDENT*, going beyond the instrument's database to tentatively identify compounds were used. *ChemSpider* is a web-based database of small molecules, with associated data (Royal Society of Chemistry, 2017). *FOR-IDENT* incorporates a database in which water relevant organic molecules, their transformation products and metabolites are listed (Grosse and Letzel, 2017). In addition, *MetFrag* was used for *in silico* fragmentation to decrease the number of potential chemical structures. The freely available software *MetFrag* annotates high precision tandem mass spectra. Candidate molecules of different databases are fragmented *in silico* and matched against *m/z* ratios of fragment spectra (Aceña et al., 2015; Ruttkies et al., 2016).

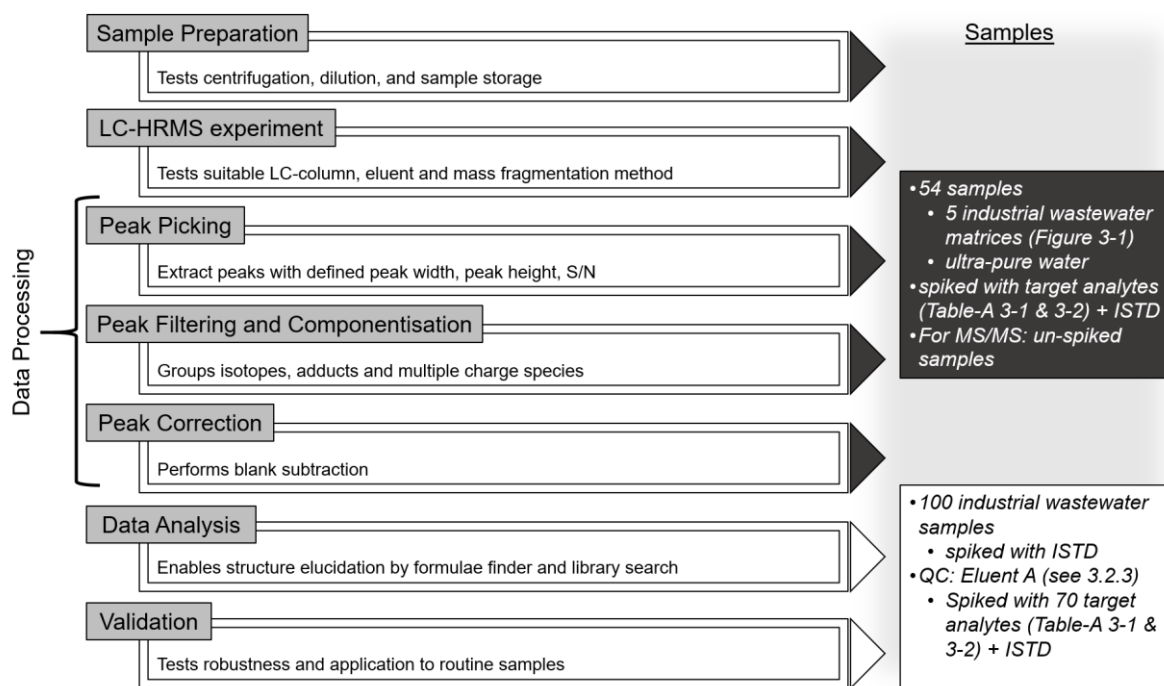


Figure 3-2 – Summary of the LC-HRMS method development including data processing and corresponding validation. The samples used for single-steps are mentioned. The performed tests are listed.

### 3.3 Results and Discussion

#### 3.3.1 Method Development and Optimisation

##### 3.3.1.1 Handling of Sample Preparation

As industrial wastewater samples can be highly loaded with matrix interferences, the sample preparation techniques centrifugation and dilution were tested regarding reducing matrix effects. For wastewater analysis, centrifugation is often applied (Mechelke et al., 2019). In comparison to filtration, it offers the opportunity of less contamination, sorption and effort. For NTS, these advantages are essential by reducing the loss of unknown compounds. Nürnberg et al. (2015) determined the recovery of intensity during the suspected-target analysis of internal standards in a diluted influent sample, which was significantly higher than in the undiluted influent sample. The overall intensities were less reduced regarding the degrees of dilution due to reduced matrix effects (Nürnberg et al., 2015). In the presented work, during centrifugation, the bulk of the particulates were removed. The experiments resulted in adequate preparation for all tested WWTP samples (from sampling sites 3 to 5, in total 45, see also Figure 3-1). All suspended particles were sedimented and all spiked target analytes were recorded during analysis (see Figure 3-3 and Table-A 3-6).

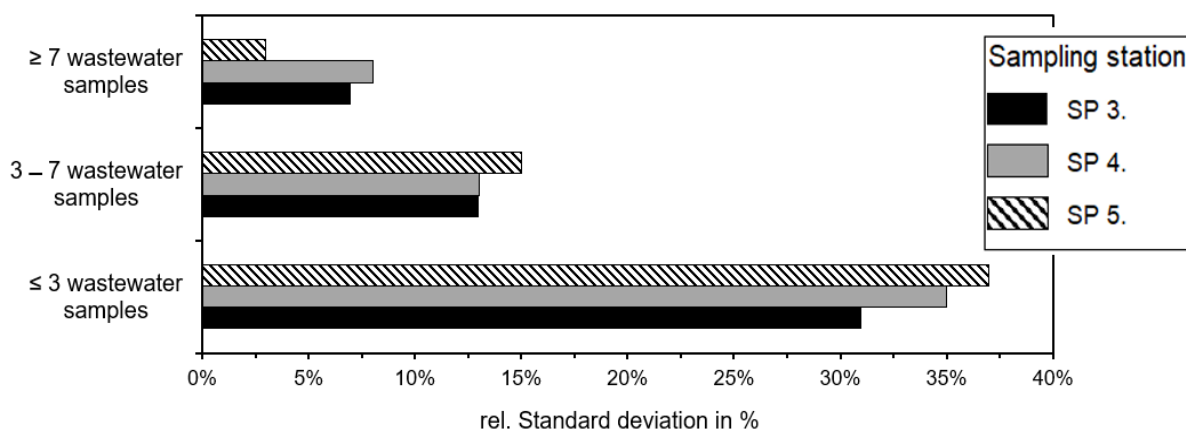


Figure 3-3 – Result of centrifugation experiments of 17 inflow samples of the WWTP (SP 3), 12 pre-treated samples within the WWTP (SP 4) and 16 samples of outflow of the WWTP (SP 5). The detailed results are shown in Appendix Table-A 3-6. The relative standard deviations (RSD) shown, describe the difference between an initial quantification measurement of spiked samples without centrifugation and the result of suspected-target screening analysis of the centrifuged samples.

For all matrices, most of the sample recovery possessed a relative standard deviation (RSD) below 10%. Only less than three samples for each sampling site have RSD of more than 30% (see Figure 3-3) compared to initial quantification measurement. However, more than half of all measured samples showed an RSD of less than 15%. Consequently, all samples were

centrifuged before injection, avoiding clogging of the LC column. Besides, sample dilution was tested to reduce the amount of loaded matrix onto the chromatographic column. In literature, it has been shown diluting influent and effluent wastewater samples reduce the matrix effects of peak enhancement or peak suppression but can lead to the loss of compounds because of insufficient sensitivity (Nürenberg et al., 2015). The saturation of the detector was assigned by evaluating the dynamic ion transmission control (ITC). The ITC modulates the ion current through the TOF-MS instrument by reducing the time ions being transmitted in each cycle, thus, protecting the detector against saturation. If modulation is significant (i.e. 0.002), the injection of the sample must be reduced. For the measured influent samples, the ITC was modulated significantly, showing saturation of the detector. Compared to the influent sample, the pre-treated wastewater samples showed more modulations for the sample without dilution than with dilution. However, the ITC modulation was more sustainable. Nevertheless, despite sample dilution, sufficient sensitivity was achieved, and a loss of compounds was not noticeable. Therefore, before analysis, the wastewater samples were diluted by a factor of 10 with mobile phase A (see 3.2.5). For a detailed description of ITC, see Appendix 3.6.2.2..

#### 3.3.1.2 *Stability Analysis during Sample Storage*

The effects of storage time and temperature were investigated by storing the samples for one week in parallel in the fridge at 8 °C and the freezer at -22 °C. During storage at 8 °C, substantially more analytes got lost or were reduced in concentration (cf. Figure 3-4 and Appendix 3.6.2.3, Table-A 3-7). In some cases, even entire losses were observed, e.g. benzenesulfonic acid. In comparison, the storage of samples for one week in the freezer at -22 °C resulted in no loss (or negligible small) of spiked analytes. The results were surprisingly in contradiction to literature (Fedorova et al., 2014), where during short-term storage, more pharmaceuticals and personal care products remained stable at 4 °C than at -18 °C. Fedorova et al. (2014) have measured other suspected-target substances in their study, which is why the stability of analytes in wastewater samples seems to be substance-specific. Furthermore, differences between the WWTP influent and effluent were observed. The WWTP effluent was less affected by degradation than influent wastewater samples (cf. Figure 3-4 and Appendix 3.6.2.3, Table-A 3-7). Comparable results were gained by Fedorova et al. (2014). Fedorova et al. analysed treated and untreated wastewater samples of a municipal WWTP, noting that stability is higher in the WWTP effluent than in influent samples (Fedorova et al., 2014). Consequently, for routine analysis, the samples are measured immediately or stored directly after centrifugation into the freezer (-22 °C) for later analysis. Otherwise, false interpretations due to losses of unknown substances during freezing will result. Nevertheless, the stability of target compounds in samples should be checked under the planned storage conditions before starting any experiment.

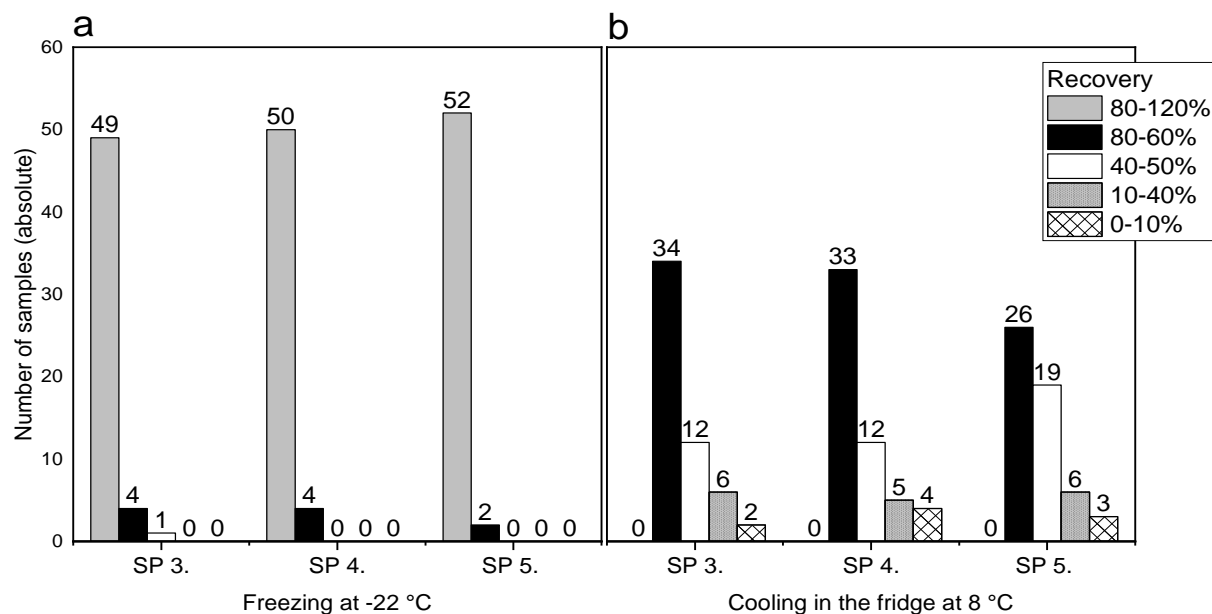


Figure 3-4 – Recovery results of storage experiments. The samples of three exemplary matrices: WWTP influent samples (SP 3), samples within the WWTP (SP 4) and WWTP effluents (SP 5) were stored for one week in the freezer at -22 °C (a) and in the fridge at 8 °C (b). The histograms show the number of analytes recovered with 80 to 120%, 60 to 80%, 40 to 50%, 10 to 40% and less than 10%. The label of the bars indicates the number of samples. The results represent means of nine measurements per matrix. Detailed results are shown in 3.6.2.3, Table-A 3-7.

Furthermore, the long-term stability of the storage in the freezer (after six months and after 14 months, see Figure-A 3-4) was determined. All analytes were reduced during storage, but the overall loss even after one year was smaller during one-week storage in the fridge. Therefore, long-term storage should be avoided and only applied in exceptional cases in combination with used ISTDs.

### 3.3.1.3 Liquid Chromatographic Conditions

For effective chromatographic separation, a combination of the mobile phase in interaction with the appropriate stationary phase was tested. C18 columns are the most often used LC columns in non-target water analysis methods (Bader et al., 2016; Jewell et al., 2019; Mechelke et al., 2019; Schymanski et al., 2014). Three polar modified and one functionally bonded LC columns were tested during method development (for a detailed description see Table 3-1). For the column tests, based on Nürenberg et al. (2015), the evaluation of suitability was determined by the following parameters: (1) recovery of all spiked target analytes and ISTD, (2) adequate selectivity, determined by retention factor (k), (3) peak shape, determined by the asymmetry factor, (4) peak width at full width at half maximum (FWHM  $\hat{=}$  width at 50%) and (5) stability of retention times determined by the relative retention times (Nürenberg et al., 2015). All parameters were evaluated over three injections of spiked analytes. The results are presented in the Appendix of Table-A 3-9 to Table-A 3-13.

Table 3-2 – Results of column comparison based on suspected-target analysis of a triplicate injection of  $10 \mu\text{g L}^{-1}$  standard for the four different columns. The following parameters were compared: retention time factor  $k$ , asymmetry factor  $AF$ , the full width at half maximum (FWHM) and standard deviation (SD) of the relative retention time (RRT). The same method was used for all four columns. A high  $k$ -value indicates that the analyte is highly retained and has spent a significant amount of time interacting with the stationary phase,  $AF$  near 1 indicates symmetric peak ( $AF > 1 \triangleq$  tailing,  $AF < 1 \triangleq$  fronting), small FWHM represents sharp peaks and small standard deviations of RRT shows stable chromatographic separation conditions.

	<b>Kinetex</b>	<b>Synergi</b>	<b>Raptor Biphenyl</b>	<b>Ultra Aqueous</b>
<b>Median of <math>k</math></b>	5.80	6.80	6.70	5.80
<b>Median <math>AF</math></b>	1.38	1.45	1.27	1.12
<b>Median FWHM</b>	0.04	0.05	0.04	0.05
<b>Median of SD of RRT</b>	0.05	0.05	0.04	0.02

Best results, regarding sensitivity and selectivity of target analytes, were observed for the Ultra Aqueous LC column in combination with a mobile phase of ultra-pure water and methanol, both spiked with 0.1% formic acid. The standard deviation for the Ultra Aqueous LC column was the lowest compared to the other tested LC columns. Additionally, the polar modified C18 phase enabled strong retention of slightly nonpolar compounds and, at the same time, selective separation of polar analytes. The asymmetry factor (see Table-A 3-11), the FWHM (see Table-A 3-12) and the standard deviation of relative retention time (see Table-A 3-13) was convincing for the Ultra Aqueous LC column in combination with the mobile phase of ultra-pure water and methanol, both spiked with 0.1% formic acid. Furthermore, the median of the asymmetry factor of  $AF = 1.12$  indicated most symmetric peaks for spiked analytes, the median of the standard deviation of relative retention time of  $SD = 0.02$  described most stable chromatographic separation and the median of the FWHM was comparable to the other columns. The retention time factor (capacity) of the Ultra Aqueous LC column (cf. Table-A 3-8) was not the highest but comparable to the other LC columns.

The mobile phase was tested for each column. Methanol was found superior to acetonitrile regarding improvement in ionisation and retention behaviour. Despite the minor decrease in mass spectrometric sensitivity in negative ionisation mode, formic acid was added to both the aqueous and organic phase as an additive for ionisation. This choice was consistent with applications (Albergamo et al., 2019; Kiefer et al., 2019; Schymanski et al., 2014). Most NTS approaches use the addition of formic acid in eluents for both positive and negative ionisation mode (Ccanccapa-Cartagena et al., 2019; Hinnenkamp et al., 2019; Hug et al., 2014; Ruttkies et al., 2019). Ammonium formate found to be a disadvantage for NTS hindering data processing because of the increased formation of adducts during ESI ionisation.

#### 3.3.1.4 Fragmentation Method in Mass Spectrometry

For identification, chemical structures are elucidated by MS fragmentation, chemical database searches and eventually confirmed via authentic standards. Therefore, in NTS full scan MS is combined with MS/MS (Gago-Ferrero et al., 2015; Mechelke et al., 2019; Schollée et al., 2015). For LC-HR-MS/MS analysis, data-dependent (DDA) and variable data-independent acquisitions (vDIA) were tested. DDA includes a full scan MS followed by MS/MS acquisition. During the survey scan, precursor ions are automatically selected above a pre-set abundance threshold triggering the instrument to perform fragmentation on those. vDIA subjects all the ions within an  $m/z$  window to fragmentation instead of selecting parent ions. For DDA and vDIA, the settings were optimised using spiked target compounds producing MS/MS spectra (see Table-A 3-1 and Table-A 3-4). In literature, vDIA fragmentation is the standard method because of generating MS/MS for all precursor ions in the untargeted analysis (Bauer et al., 2019; Périat et al., 2019; Wang et al., 2019; Zhu et al., 2014). Nevertheless, in this study the DDA fragmentation method was found to be more suitable for reliable, high-quality MS/MS spectra. Following the development of an optimised vDIA method for wastewater analysis, DDA and vDIA were compared in terms of reproducibility and sensitivity. In total, 47, 46 and 46 suspected-targets and ISTDs were identified from a triplicate DDA analysis of authentic industrial wastewater (SP 3, see Figure 3-1), with 43 analytes (51.2%) found in all three triplicates. The corresponding analysis applying the vDIA method with variable windows resulted in the identification of 46, 43 and 40 in the wastewater, with a total of 39 suspected-targets and ISTDs (46.4%) found in all three replicates (cf. Table-A 3-14). Furthermore, the intensities of the identified TrOCs were smaller for vDIA compared to DDA, demonstrating the sensitivity of the DIA methodology. Comparing both fragmentation methods for spiked samples, vDIA shows a higher baseline (noise), fewer data points per peak, smaller fragment ions and difficulties in interpretation of MS/MS spectra because of many fragment ions. The acquired fragmentation spectra were challenging to analyse due to the lack of parent ion selection, especially for wastewater samples. Furthermore, some signals of DDA MS/MS spectra did not appear in vDIA spectra or were asymmetric. Nevertheless, in further studies, using spiked wastewater samples, for both vDIA and DDA, an MS/MS spectrum was observed for all spiked target analytes down to a concentration of  $5 \mu\text{g L}^{-1}$  (using spiked wastewater sample). However, vDIA had no increase in information compared to DDA. MS/MS spectra of vDIA were either comparable in quality or superimposable in fragmentation to DDA, especially working with wastewater matrix, resulting in no evaluable spectra. Sometimes, in vDIA, there occur more fragment ions comparing to DDA, which indicates the difficulty in the classification of fragment ions back to precursor ions. Overall, DDA-based methods acquired, in agreement with other studies, qualitatively better MS/MS spectra but with a lower MS/MS acquisition hit

rate than DIA (Barbier et al., 2020; Zhu et al., 2014). For example, Zhu et al. (2014) determined DDA fragmentation methods produce a higher quality of MS/MS compared to DIA methods which are necessary for successful identification of TrOC.

### 3.3.2 Data Processing

The flowchart of data processing is shown in Figure 3-2. The starting point of data processing was a set of raw data files consisting of successively recorded MS full scan spectra. The data processing strategy for NTS was controlled and validated by measuring spiked authentic wastewater samples evaluated by a suspect list (70 compounds, see Table-A 3-1). Table 3-3 shows the results of the evaluation of spiked authentic wastewater samples of four exemplary matrices (SP 2 – SP 5). For evaluation, peak picking was assessed for the quality parameters of Table 3-3 (true positive, true negative, false positive and false negative). The results were categorised to the listed quality parameter by determination of the true peak by visually evaluated the peak shape (symmetric peak with almost Gaussian function is favoured) and sufficient differentiation to baseline (assigned by the signal-to-noise ratio of  $S/N \geq 10$ ). Measuring spiked authentic wastewater samples (174 in positive ionisation mode and 108 in negative ionisation mode), most false positives were identified based on non-matching retention times or high mass errors when compared to the authentic standard. Both ionisation modes (ESI<sup>+</sup> and ESI<sup>-</sup>) were considered. Retention time shifts could occur due to sample-solvent induced matrix effects on LC behaviour observed previously (Fang et al., 2015).

*Table 3-3 – Result of unknown screening of authentic wastewater samples. True positives/negatives (TP / TN), false positives (FP) and false negatives (FN) were assigned to get an impression of the suitability of the non-target method. The quality parameters are given in % and are assigned based on suspected-target list matches of treated (SP 5), pre-treated (SP 4) and untreated (SP 2, SP 3) wastewater samples. Overall, 174 samples were measured in positive ionisation mode and 108 were measured in negative ionisation mode.*

<b>N(+)</b> = 174;	<b>N(-)</b> = 108	<b>TP/TN</b>	<b>FP</b>	<b>FN</b>
<b>SP 2</b>	(+)	93.2%	3.42%	3.40%
	(-)	81.5%	4.62%	13.9%
<b>SP 3</b>	(+)	94.0%	4.31%	1.72%
	(-)	91.7%	4.63%	3.70%
<b>SP 4</b>	(+)	94.9%	3.42%	1.71%
	(-)	88.0%	4.62%	7.41%
<b>SP 5</b>	(+)	89.7%	6.86%	3.42%
	(-)	92.6%	0.920%	6.48%
<b>Mean</b>	(+), (-)	90.7%	4.10%	5.20%

The virtually controlled peak picking showed true peaks were identified for features with a defined peak width of 6 points, a set peak height of 100 cps (XIC) and a retention time window of 1 to 26 min (neglecting the salt load of the sample; parameters are listed in the Appendix

Table-A 3-3). Besides, using an intensity threshold value of 100 cps, all spiked target components were recovered (see Table-A 3-15). An increased threshold decreased the number of recovered targets (cf. Table-A 3-15 to Table-A 3-17). For 500 cps 8.48% of targets were not detected and for 1,000 cps 19.7%. In comparison, the number of false positives increases for lower intensity thresholds. False positives were described by hits due to improper integration of the chromatographic background, injection peaks or matrix signals as peaks. Nevertheless, this effect is comparably weak instead of false negative results, i.e. the peak finding algorithm missed a real peak of interest. Therefore, the intensity threshold of 100 cps was applied, which, on the one hand, favours true positives and on the other side, reduces false negatives.

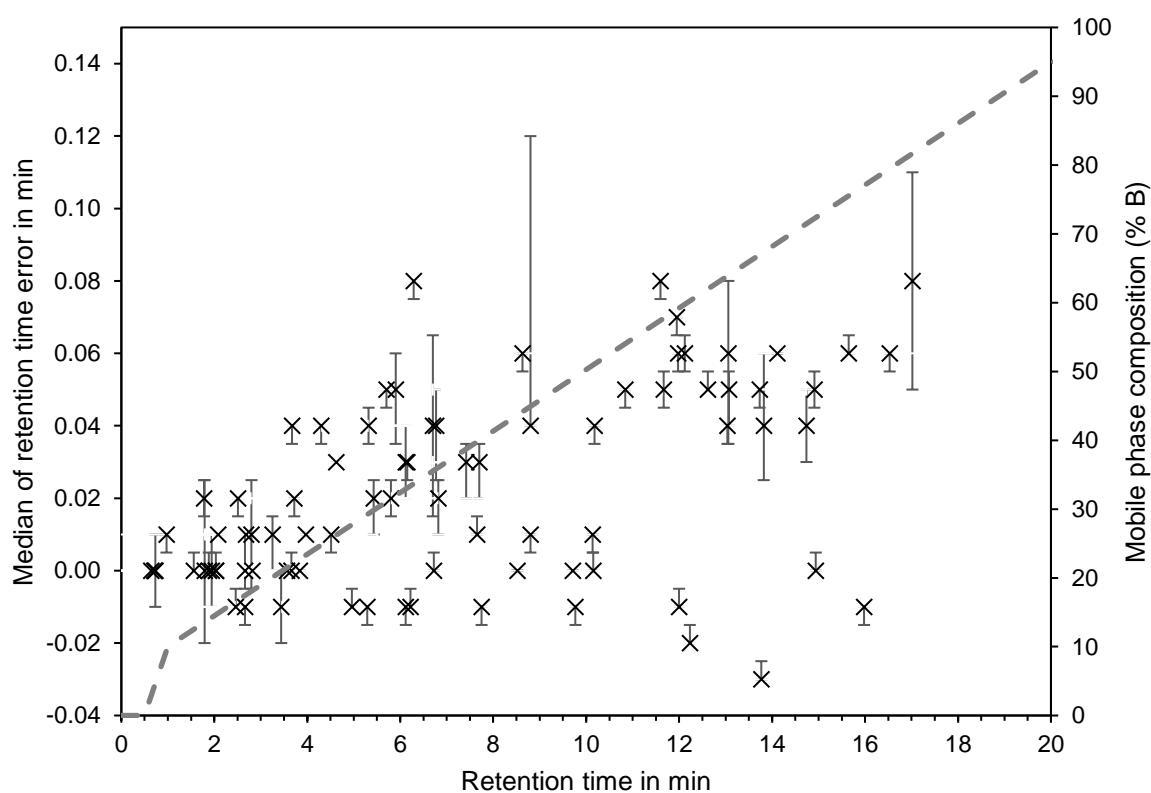


Figure 3-5 – Median of retention time deviation in min of triplicate measurement of all spiked suspected-targets and ISTDs over absolute retention time in min. The error bars represent the upper (maximum of the third quartile, 75<sup>th</sup> percentile) and lower (minimum of first quartile, 25<sup>th</sup> percentile) whiskers with respectively N = 3. Detailed results are listed in 3.6.2.8, Table-A 3-18.

Moreover, the results showed stable conditions of retention times. However, the retention time tolerance was specified, because significant retention time shifts would result in an impossible interpretation of unknown substances and, consequently, favours false-negatives. For measured targets, a maximal retention time error of 0.08 min  $\pm$  0.03 was achieved (see Figure 3-5, Table-A 3-18). Therefore, in NTS, for identification criteria, an error of  $\pm$  0.1 min was applied.



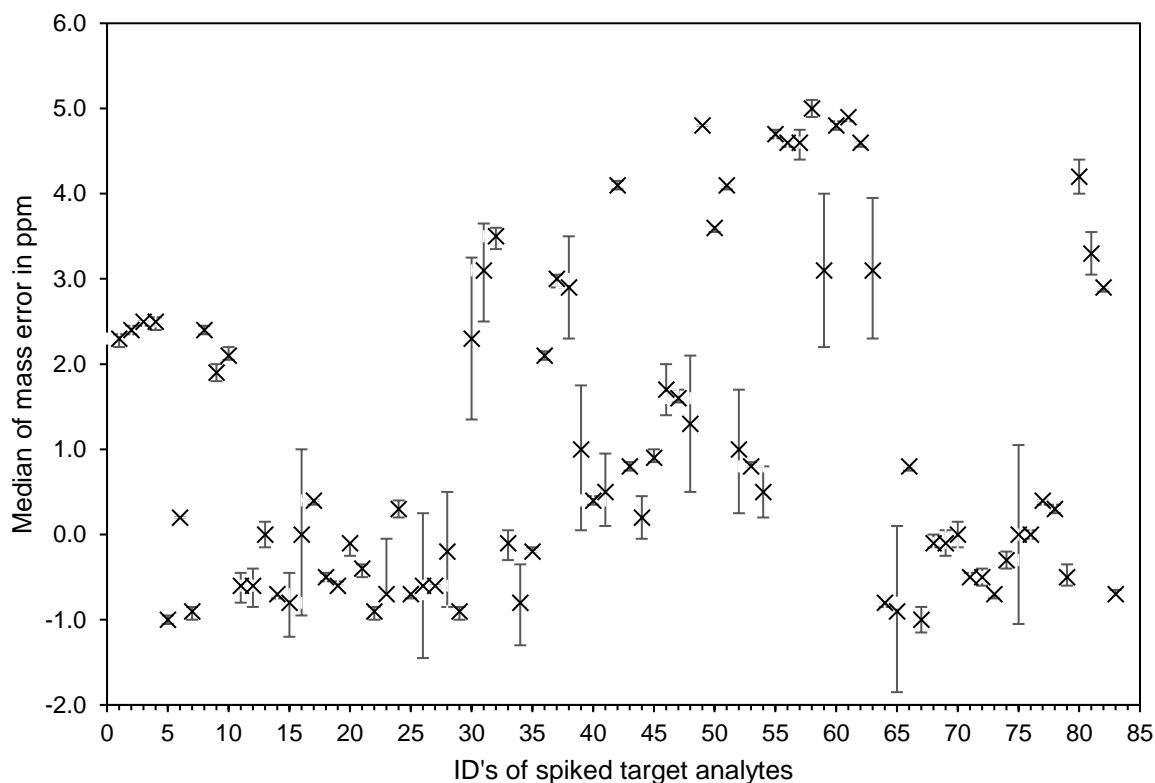


Figure 3-6 –The figure illustrates the median of mass error of spiked suspected-targets and ISTDs in ppm of all measured samples for validation in positive and negative ionisation mode over three months. The error bars represent the upper (maximum of the third quartile, 75<sup>th</sup> percentile) and lower (minimum of first quartile, 25<sup>th</sup> percentile) whiskers. Detailed results are listed in 3.6.2.8, Table-A 3-18.

Besides retention time tolerance, the mass error tolerance was specified to check the stability of the used data processing method. Figure 3-6 shows the mass error of all spiked suspected-targets and ISTDs over all measured samples. Mass errors are caused by complex matrices, presence of isobaric interferences, a saturation of the detector or by the instrument itself (Kaufmann and Butcher, 2007; Kaufmann and Walker, 2013). For positive ionisation mode, the mass error over all measured suspected-targets ranged from  $-1.11$  ppm to  $4.91$  ppm and for negative ionisation mode from  $-0.05$  ppm to  $3.95$  ppm. These errors showed the high stability of the mass spectrometer and, thus, allow the use of the narrow mass extraction window of  $5$  ppm, which corroborates results of previous studies showing reproducible results with a  $5$  ppm window (Chitescu et al., 2015; Kiefer et al., 2019; Mechelke et al., 2019; Verkh et al., 2018).

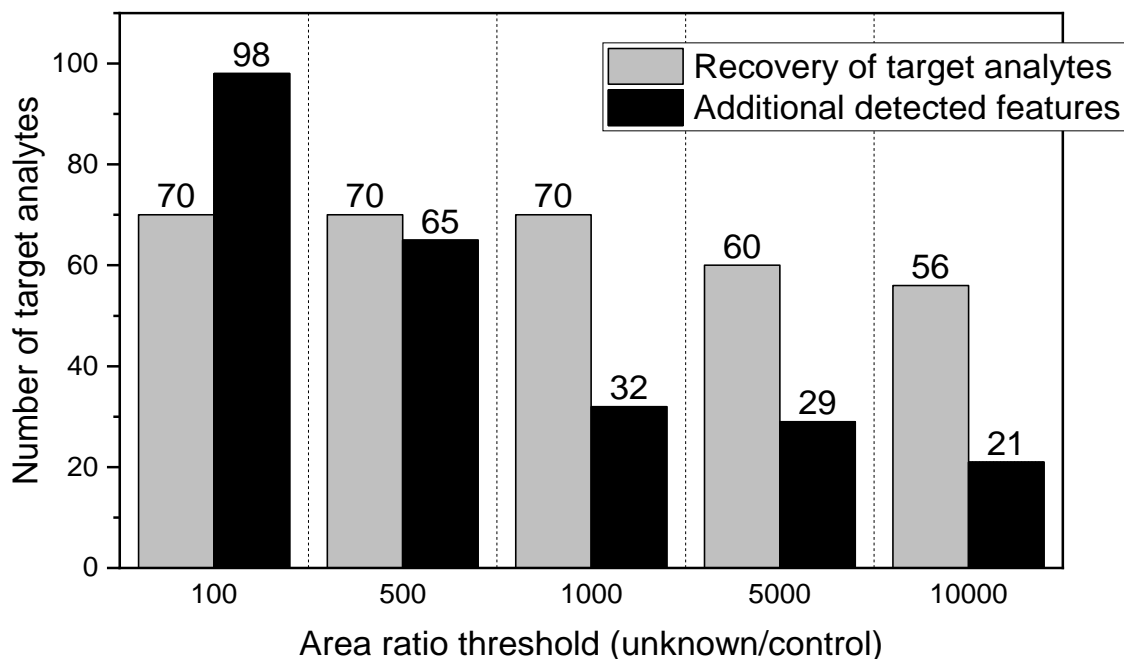


Figure 3-7 – Recovery of target analytes and additionally detected features in blank using different area ratio thresholds for blank subtraction.

Afterwards, the detected features were extracted, smoothed and integrated according to Table-A 3-3 b in the Appendix. Signals from isotopic peaks, adducts, different charge states and in-source fragment ions of a compound were grouped in componentisation. For SCIEX OS the componentisation was performed automatically. An automated blank correction was performed to reduce the complexity of the sample and eliminate features, occurring from the LC-MS system or other contamination, for example from the solvents (Weber et al., 2012). For blank subtraction, a solvent blank was processed identically to the samples. In Table-A 3-3 e) the parameter for blank subtraction is listed. Identifying the best threshold ratio for the relation of unknown (wastewater sample) to control (blank), the target analytes were spiked into the blank sample and were evaluated. Both the recovery of target analytes and the additional detection of other features were considered in the evaluation of the results. For peak picking the best results, based on the highest recovery of spiked validation compounds and comparably less detection of other features, was obtained for a 1000:1 intensity ratio in the sample compared to the blank sample (see Figure 3-7).

### 3.3.3 Validation of Non-Target Screening based on developed LC-HRMS

For validation, 100 un-spiked samples were used to test the suitability of the method. The samples were selected on a previous analysis (using the routine analysis of LC-UV). Samples which could be assigned positive for at least one of the targets of Table-A 3-1 in the Appendix were used for validation. The results of the suspected-target analysis are listed in the Appendix

(see Table-A 3-19). The peak table in Table-A 3-20 is the result of the merged features extracted with *SCIEX OS 'non-target analysis'* tool of a wastewater sample of the WWTP outflow (SP 5, see Figure 3-1). Further interpretations of the resulted feature-list based upon identification tasks. Tentative identification of compounds in NTS required additional manual processing and the aid of an on-line database (e.g. ChemSpider, FOR-IDENT and *MetFrag*) going beyond the instrument's database. However, for such a large number of processed features of up to 400 (see 1.6.2.10; Table-A 3-20), the identification process is an unreasonable demand. Therefore, prioritisation techniques are required. Features are prioritised, e.g. based on their intensity (Aceña et al., 2015; Ruff et al., 2015; Samanipour et al., 2019). The changes in signal heights are theoretically mirrored by alterations in concentrations. Matrix effects such as ion suppression may affect this relation. Therefore, candidates are also prioritised by their structure information, e.g. specific isotopic patterns (Gago-Ferrero et al., 2015), their toxicity (Guo et al., 2019) or frequency of appearance during time series analysis (see Chapter 4). A detailed description of an identification workflow in industrial wastewater is presented in chapter six of this dissertation.

### 3.4 Conclusion

For industrial wastewater samples, a new analysis method was developed, which expands the monitoring of wastewater on LC-HRMS (qTOF) analysis. The Ultra Aqueous LC column proved to be suitable, separating TrOC of medium polar to polar character during a multicomponent screening approach. For MS/MS analysis going beyond the MS full scan method, DDA proved to be suitable for the acquisition of high-quality MS/MS spectra. The suitability of the method was tested for 70 target substances in industrial wastewater samples. Besides, a proper data evaluation was found, including peak picking, specific filtering and blank subtraction algorithms. For the measured suspected-targets, the retention time and the *m/z* ratio were stable among analysis sequences and the whole measurement workflow of several months. The use of blank subtraction in combination with stringent filter criteria led to an enhanced data quality since the peak recognition could be improved considerably. Blank subtraction led to a simplified data evaluation since interferences occurring during analysis or arise from the system were reduced. Nevertheless, isotope-labelled internal standards and quality control samples are necessary to control the evaluation method within each batch. Finally, the validation of the data processing strategy demonstrated the suitability of the unknown screening for industrial wastewater. The application to authentic industrial wastewater samples was shown. Besides, differences in the stability of effluent and influent samples were observed, which will require special care in evaluating the efficiency of wastewater treatment in the future.

### 3.5 References

- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography - High-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 407, 6289–6299. <https://doi.org/10.1007/s00216-015-8852-6>
- Albergamo, V., Schollée, J.E., Schymanski, E.L., Helmus, R., Timmer, H., Hollender, J., de Voogt, P., 2019. Nontarget Screening Reveals Time Trends of Polar Micropollutants in a Riverbank Filtration System. *Environ. Sci. Technol.* 53, 7584–7594. <https://doi.org/10.1021/acs.est.9b01750>
- Appa, R., Mhaisalkar, V.A., Bafana, A., Et.al., 2018. Simultaneous quantitative monitoring of four indicator contaminants of emerging concern (CEC) in different water sources of Central India using SPE/LC-(ESI)MS-MS. *Environmental Monit. Assess.* 190. <https://doi.org/https://doi.org/10.1007/s10661-018-6867-0>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2017. LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes. *Anal. Chem.* 89, 13219–13226. <https://doi.org/10.1021/acs.analchem.7b03037>
- Bader, T., Schulz, W., Lucke, T., 2016. Application of non-target analysis with LC-HRMS for the monitoring of raw and potable water: Strategy and results, in: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 2. ACS Symposium Series, Washington, DC.*, pp. 49–70. <https://doi.org/10.1021/bk-2016-1242.ch003>
- Barbier, P., Hilaire, S., Rousseau, K., Seyer, A., Dechaumet, S., Damont, A., Junot, C., Fenaille, F., 2020. Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics. *Metabolites* 10, 1–17. <https://doi.org/doi:10.3390/metabo10040158>
- Bauer, A., Jesús, F., Gómez Ramos, M.J., Lozano, A., Fernández-Alba, A.R., 2019. Identification of unexpected chemical contaminants in baby food coming from plastic packaging migration by high resolution accurate mass spectrometry. *Food Chem.* 295, 274–288. <https://doi.org/10.1016/j.foodchem.2019.05.105>
- Bonner, R., Hopfgartner, G., 2019. SWATH data independent acquisition mass spectrometry for metabolomics. *TrAC - Trends Anal. Chem.* 120. <https://doi.org/10.1016/j.trac.2018.10.014>
- Brüggen, S., Schmitz, O.J., 2018. A New Concept for Regulatory Water Monitoring Via High-Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. *J. Anal. Test.* 2, 342–351. <https://doi.org/10.1007/s41664-018-0081-5>
- Ccancapa-Cartagena, A., Pico, Y., Ortiz, X., Reiner, E.J., 2019. Suspect, non-target and target screening of emerging pollutants using data independent acquisition: Assessment of a Mediterranean River basin. *Sci. Total Environ.* 687, 355–368. <https://doi.org/10.1016/j.scitotenv.2019.06.057>
- Chitescu, C.L., Kaklamanos, G., Nicolau, A.I., Stolker, A.A.M.L., 2015. High sensitive multiresidue analysis of pharmaceuticals and antifungals in surface water using U-HPLC-Q-Exactive Orbitrap HRMS. Application to the Danube river basin on the Romanian territory. *Sci. Total Environ.* 532, 501–511. <https://doi.org/10.1016/j.scitotenv.2015.06.010>
- Fang, N., Yu, S., Ronis, M.J.J., Badger, T.M., 2015. Matrix effects break the LC behavior rule for analytes in LC-MS/MS analysis of biological samples. *Exp. Biol. Med.* 240, 488–497. <https://doi.org/10.1177/1535370214554545>

- Fedorova, G., Golovko, O., Randak, T., Grabic, R., 2014. Storage effect on the analysis of pharmaceuticals and personal care products in wastewater. *Chemosphere* 111, 55–60. <https://doi.org/10.1016/j.chemosphere.2014.02.067>
- Fiorino, G.M., Fresch, M., Brümmer, I., Losito, I., Arlorio, M., Brockmeyer, J., Monaci, L., 2019. Mass Spectrometry-Based Untargeted Proteomics for the Assessment of Food Authenticity: The Case of Farmed Versus Wild-Type Salmon. *J. AOAC Int.* 1–7. <https://doi.org/10.5740/jaoacint.19-0062>
- Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* 49, 12333–12341. <https://doi.org/10.1021/acs.est.5b03454>
- Geppert, H., Vogt, M., Bajorath, J., 2010. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model.* 50, 205–216. <https://doi.org/10.1021/ci900419k>
- Grosse, S., Letzel, T., 2017. User Manual for FOR-IDENT Database 3, 1–35.
- Guo, J., Deng, D., Wang, Y., Yu, H., Shi, W., 2019. Extended suspect screening strategy to identify characteristic toxicants in the discharge of a chemical industrial park based on toxicity to *Daphnia magna*. *Sci. Total Environ.* 650, 10–17. <https://doi.org/10.1016/j.scitotenv.2018.08.215>
- Hinnenkamp, V., Balsaa, P., Schmidt, T.C., 2019. Quantitative screening and prioritization based on UPLC-IM-Q-TOF-MS as an alternative water sample monitoring strategy. *Anal. Bioanal. Chem.* <https://doi.org/https://doi.org/10.1007/s00216-019-01994-w>
- Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.* 92, 1898–1907. <https://doi.org/https://doi.org/10.1021/acs.analchem.9b04095>
- Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* 51, 11505–11512. <https://doi.org/10.1021/acs.est.7b02184>
- Horlacher, O., Lisacek, F., Müller, M., 2016. Mining Large Scale Tandem Mass Spectrometry Data for Protein Modifications Using Spectral Libraries. *J. Proteome Res.* 15, 721–731. <https://doi.org/10.1021/acs.jproteome.5b00877>
- Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* 184, 25–32. <https://doi.org/10.1016/j.envpol.2013.07.048>
- Jewell, K.S., Kunkel, U., Ehlig, B., Thron, F., Schlüsener, M., Dietrich, C., Wick, A., Ternes, T.A., 2019. Comparing mass, retention time and MS<sup>2</sup> spectra as criteria for the automated screening of small molecules in aqueous environmental samples analyzed by LC-QToF-MS/MS. *Rapid Commun. Mass Spectrom.* rcm.8541. <https://doi.org/10.1002/rcm.8541>
- Katajamaa, M., Orešič, M., 2007. Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* 1158, 318–328. <https://doi.org/10.1016/j.chroma.2007.04.021>
- Kaufmann, A., Butcher, P., 2007. Strategies to avoid false negative findings in residue analysis using liquid chromatography coupled to time-of-flight mass spectrometry. *Anton. Rapid Commun. Mass Spectrom.* 20, 4065–4072. <https://doi.org/10.1002/rcm>

- Kaufmann, A., Walker, S., 2013. Evaluation of the interrelationship between mass resolving power and mass error tolerances for targeted bioanalysis using liquid chromatography coupled to high-resolution mass spectrometry. *Rapid Commun. Mass Spectrom.* 27, 347–356. <https://doi.org/10.1002/rcm.6454>
- Kiefer, K., Müller, A., Singer, H., Hollender, J., 2019. New relevant pesticide transformation products in groundwater detected using target and suspect screening for agricultural and urban micropollutants with LC-HRMS. *Water Res.* 165, 114972. <https://doi.org/10.1016/j.watres.2019.114972>
- Köppe, T., Jewell, K.S., Dietrich, C., Wick, A., Ternes, T.A., 2020. Application of a non-target workflow for the identification of specific contaminants using the example of the Nidda river basin. *Water Res.* 178. <https://doi.org/10.1016/j.watres.2020.115703>
- Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–51. <https://doi.org/10.1007/s00216-010-3608-9>
- Marta-Sanchez, A.V., Caldas, S.S., Schneide, A., Et.al., 2018. Trace analysis of parabens preservatives in drinking water treatment sludge, treated, and mineral water samples. *Environ. Sci. Pollut. Res.* 25, 14460–14470. <https://doi.org/https://doi.org/10.1007/s11356-018-1583-4>
- Mechelke, J., Longrée, P., Singer, H., Hollender, J., 2019. Vacuum-assisted evaporative concentration combined with LC-HRMS/MS for ultra-trace-level screening of organic micropollutants in environmental water samples. *Anal. Bioanal. Chem.* 2555–2567. <https://doi.org/10.1007/s00216-019-01696-3>
- Nürenberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography - high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>
- Park, N., Choi, Y., Kim, D., Kim, K., Jeon, J., 2018. Prioritization of highly exposable pharmaceuticals via a suspect/non-target screening approach: A case study for Yeongsan River, Korea. *Sci. Total Environ.* 639, 570–579. <https://doi.org/10.1016/j.scitotenv.2018.05.081>
- Pérez-Fernández, V., Mainero Rocca, L., Tomai, P., Fanali, S., Gentili, A., 2017. Recent advancements and future trends in environmental analysis: Sample preparation, liquid chromatography and mass spectrometry. *Anal. Chim. Acta* 983, 9–41. <https://doi.org/10.1016/j.aca.2017.06.029>
- Périat, A., Bieri, S., Mottier, N., 2019. SWATH-MS screening strategy for the determination of food dyes in spices by UHPLC-HRMS. *Food Chem. X* 1, 100009. <https://doi.org/10.1016/j.fochx.2019.100009>
- Pozo, Ó.J., Sancho, J. V., Ibáñez, M., Hernández, F., Niessen, W.M.A., 2006. Confirmation of organic micropollutants detected in environmental samples by liquid chromatography tandem mass spectrometry: Achievements and pitfalls. *TrAC - Trends Anal. Chem.* 25, 1030–1042. <https://doi.org/10.1016/j.trac.2006.06.012>
- Richardson, S.D., Ternes, T.A., 2018. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* 90, 398–428. <https://doi.org/10.1021/acs.analchem.7b04577>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Ruff, M., Mueller, M.S., Loos, M., Singer, H.P., 2015. Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-

- resolution mass-spectrometry - Identification of unknown sources and compounds. *Water Res.* 87, 145–154. <https://doi.org/10.1016/j.watres.2015.09.017>
- Ruttkies, C., Schymanski, E.L., Williams, A.J., Krauss, M., 2019. Supporting non-target identification by adding hydrogen deuterium exchange MS / MS capabilities to MetFrag. *Anal. Bioanal. Chem.* 411, 4683–4700. <https://doi.org/https://doi.org/10.1007/s00216-019-01885-0>
- Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 1–16. <https://doi.org/10.1186/s13321-016-0115-9>
- Samanipour, S., O'Brien, J.W., Reid, M., Thomas, K. V., 2019. A Self Adjusting Algorithm for the Non-targeted Feature Detection of High Resolution Mass Spectrometry Coupled with Liquid Chromatography Profile Data. *Anal. Chem.* [acs.analchem.9b02422](https://doi.org/10.1021/acs.analchem.9b02422). <https://doi.org/10.1021/acs.analchem.9b02422>
- Saurina, J., Sentellas, S., 2019. Liquid chromatography coupled to mass spectrometry for metabolite profiling in the field of drug discovery. *Expert Opin. Drug Discov.* 14, 469–483. <https://doi.org/10.1080/17460441.2019.1582638>
- Schmidt, T.C., 2018. Recent trends in water analysis triggering future monitoring of organic micropollutants. *Anal. Bioanal. Chem.* 410, 3933–3941. <https://doi.org/10.1007/s00216-018-1015-9>
- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* 87, 12121–12129. <https://doi.org/10.1021/acs.analchem.5b02905>
- Schulz, W., Seitz, W., Lucke, T., 2014. Non-Target-Analytik deckt neue Spurenstoffe auf [WWW Document]. Aktuelle Wochenschau der GDCh. URL <http://www.aktuelle-wochenschau.de/main-navi/archiv/wasserchemie-2014/kw44-non-target-analytik-deckt-neue-spurenstoffe-auf.html> (accessed 5.29.18).
- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014. Strategies to characterize polar organic contamination in wastewater: Exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol.* 48, 1811–1818. <https://doi.org/10.1021/es4044374>
- Verkh, Y., Rozman, M., Petrovic, M., 2018. A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. *Chemosphere* 200, 397–404. <https://doi.org/10.1016/j.chemosphere.2018.02.095>
- Wang, R., Yin, Y., Zhu, Z.-J., 2019. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Anal. Bioanal. Chem.* 1–9. <https://doi.org/10.1007/s00216-019-01709-1>
- Weber, R.J.M., Eva, L., Bruty, J., He, S., Viant, M.R., 2012. MaConDa: a publicly accessible Mass spectrometry Contaminants Database. *Bioinformatics.* <https://doi.org/doi:10.1093/bioinformatics/bts527>.
- Zhu, X., Chen, Y., Subramanian, R., 2014. Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* 86, 1202–1209. <https://doi.org/10.1021/ac403385y>

## **3.6 Chapter Appendix**

### **3.6.1 Supplement for Chapter Materials and Method**

#### *3.6.1.1 Targets and ISTD*

In the following, the used target analytes and ISTD with corresponding mass information are listed.



Table-A 3-1 – Table of analytes and ISTD used for method development, method validation and quality control (QC). All data are rounded data. Fragment ions are listed in decreasing intensity from 1 to 5. RT  $\triangleq$  Retention time (\*Industrial chemical).

Target compounds	Sum formula	CAS No.	Main Ion / Adduct	RT in min	Expected Mass in Da	Accurate Mass in Da	1	2	3	4	5
1-(3'-Sulfophenyl)-3-methyl-5-aminopyrazole	C <sub>10</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub> S	23646-86-8	[M-H] <sup>-</sup>	2.46	252.045	252.0446	79.9578	188.0831	107.0376	-	-
1,5-Diaminonaphthalene	C <sub>10</sub> H <sub>10</sub> N <sub>2</sub>	2243-62-1	[M+H] <sup>+</sup>	1.94	159.092	159.0915	115.0540	143.0740	159.0910	130.0650	89.0380
1,6-Naphthalenedisulfonic acid	C <sub>10</sub> H <sub>6</sub> O <sub>6</sub> S <sub>2</sub>	525-37-1	[M-H] <sup>-</sup>	2.03	286.969	286.9688	207.0123	143.0509	286.9693	79.9576	-
1,7-Naphthalenedisulfonic acid	C <sub>10</sub> H <sub>6</sub> O <sub>6</sub> S <sub>2</sub>	5724-16-3	[M-H] <sup>-</sup>	2.03	286.969	286.9687	207.0124	143.0509	286.9693	79.9576	-
2-(n-Ethylanilino)ethanol	C <sub>10</sub> H <sub>16</sub> NO	92-50-2	[M+H] <sup>+</sup>	2.68	166.123	166.1223	106.0650	120.0809	77.0387	166.1224	45.0333
2-(Trifluoromethyl)benzenesulfonamide/ TBSA	C <sub>7</sub> H <sub>6</sub> F <sub>3</sub> NO <sub>2</sub> S	1869-24-5	[M-H] <sup>-</sup>	4.96	224.000	223.9997	77.9656	160.0384	100.0192	224.0000	145.0282
2-(Trifluoromethyl)benzamide	C <sub>8</sub> H <sub>6</sub> F <sub>3</sub> NO	360-64-5	[M+H] <sup>+</sup>	5.32	190.047	190.0473	130.0300	102.0341	150.0358	170.0418	75.0234
2,2'-(4-Methylphenylimino)diethanol	C <sub>11</sub> H <sub>17</sub> NO <sub>2</sub>	430227	[M+H] <sup>+</sup>	3.25	196.133	196.1328	120.0806	134.0966	91.0543	196.1337	45.0336
2,4,6-Triamino-1,3,5-triazine/ Melamine	C <sub>3</sub> H <sub>6</sub> N <sub>6</sub>	108-78-1	[M+H] <sup>+</sup>	0.73	127.073	127.0724	85.0510	43.0289	68.0246	127.0727	110.0468
2,4,6-Trimethoxy-1,3,5-triazine	C <sub>8</sub> H <sub>9</sub> N <sub>3</sub> O <sub>3</sub>	-	[M+H] <sup>+</sup>	6.77	172.072	172.0714	72.0444	58.0288	83.0241	140.0461	-
2-Amino-4-methoxy-6-trifluoromethyl-1,3,5-triazine/ AMTT	C <sub>9</sub> H <sub>5</sub> F <sub>3</sub> N <sub>4</sub> O	1245970	[M+H] <sup>+</sup>	5.90	195.049	195.0487	110.0222	43.0291	57.0447	57.0447	68.9951
2-Chlorobenzoic acid	C <sub>7</sub> H <sub>5</sub> O <sub>2</sub> Cl	118-91-2	[M+H] <sup>+</sup>	6.72	157.005	157.0050	77.0387	111.0000	138.9951	51.0233	-
2-Naphthalenesulfonic acid	C <sub>10</sub> H <sub>8</sub> SO <sub>3</sub>	120-18-3	[M-H] <sup>-</sup>	5.29	207.012	207.0120	143.0505	207.0124	79.9575	115.0557	-
3,5-Dihydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>	99-10-5	[M-H] <sup>-</sup>	2.66	153.093	153.0193	109.0299	41.0033	67.0190	65.0008	153.0244
3-Aminobenzenesulfonic acid	C <sub>6</sub> H <sub>7</sub> NO <sub>3</sub> S	121-47-1	[M-H] <sup>-</sup>	0.71	172.007	172.0072	79.9572	172.0078	108.0455	40.0193	-

Table-A 3-1 continued.

3-Hydroxy-2-nitrobenzoic acid	C <sub>7</sub> H <sub>5</sub> NO <sub>5</sub>	602-00-6	[M-H] <sup>-</sup>	3.44	182.009	182.0093	92.0266	45.9929	108.0235	138.0193	-
3-Nitrophthalic acid	C <sub>8</sub> H <sub>5</sub> NO <sub>6</sub>	603-11-2	[M-H] <sup>-</sup>	2.66	210.004	210.0041	45.9933	122.0249	163.9552	-	-
4-Aminobenzoic acid	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub>	150-13-0	[M+H] <sup>+</sup>	2.80	138.055	138.0549	77.0387	65.0386	92.0497	120.0449	51.0226
4-Chlor-2-cyano-5-(4-methylphenyl)imidazole/ CCIM	C <sub>11</sub> H <sub>8</sub> ClN <sub>3</sub>	120118-14-1	[M+H] <sup>+</sup>	13.06	218.048	218.0478	183.0790	139.0310	130.0650	103.0550	118.0660
4-Chlorobenzoic acid	C <sub>7</sub> H <sub>5</sub> ClO <sub>2</sub>	74-11-3	[M-H] <sup>-</sup>	9.77	154.991	154.9904	34.9692	111.0009	44.9978	154.9964	-
4-Hydroxyphenylglycolic acid	C <sub>8</sub> H <sub>8</sub> O <sub>4</sub>	1198-84-1	[M-H] <sup>-</sup>	1.56	167.035	167.0350	-	-	-	-	-
4-Sulphophthalic acid	C <sub>8</sub> H <sub>6</sub> O <sub>7</sub> S	89-08-07	[M-H] <sup>-</sup>	0.97	244.976	244.9767	-	-	-	-	-
4-Toluenesulfonic acid	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> S	6192-52-5	[M-H] <sup>-</sup>	2.82	171.012	171.0121	79.9571	171.0120	107.0503	-	-
6-Chloropyridine-2-carboxylic acid	C <sub>6</sub> H <sub>4</sub> ClNO <sub>2</sub>	4684-94-0	[M+H] <sup>+</sup>	4.30	158.000	158.0002	111.9948	76.0183	139.9901	84.9840	75.0106
Acifluorfen	C <sub>14</sub> H <sub>7</sub> ClF <sub>3</sub> NO <sub>5</sub>	50594-66-5	[M-H] <sup>-</sup>	13.77	359.989	359.9890	194.9836	222.0303	286.0016	315.9991	45.9938
Azoxystrobin	C <sub>22</sub> H <sub>17</sub> N <sub>3</sub> O <sub>5</sub>	131860-33-8	[M+H] <sup>+</sup>	13.07	404.124	404.1234	372.0970	344.1026	329.0792	345.1056	330.0820
Azoxystrobin-D4	C <sub>22</sub> H <sub>13</sub> D <sub>4</sub> N <sub>3</sub> O <sub>5</sub>	-	[M+H] <sup>+</sup>	13.04	408.149	408.1487	377.1275	349.1336	-	-	-
ABC 700*	C <sub>8</sub> H <sub>16</sub> F <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	-	[M+H] <sup>+</sup>	3.56	177.047	177.1234	137.0345	42.0344	109.0399	177.0489	-
Benzamideoxime	C <sub>7</sub> H <sub>8</sub> N <sub>2</sub> O	613-92-3	[M+H] <sup>+</sup>	1.78	137.071	137.0709	77.0389	104.0500	51.0233	137.0712	119.0610
Benzenesulfonic acid	C <sub>6</sub> H <sub>6</sub> O <sub>3</sub> S	98-11-3	[M-H] <sup>-</sup>	1.79	156.996	156.9963	79.9576	93.0347	156.9967	-	-
Benzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	65-85-0	[M-H] <sup>-</sup>	3.66	121.030	121.0296	77.0408	121.0266	92.0274	-	-
Bezafibrate-D6	C <sub>19</sub> H <sub>14</sub> D <sub>6</sub> ClNO <sub>4</sub>	1219802-74-0	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	12.62 / 12.61	366.138	366.1380	274.0642	154.0067	322.1486	276.0801 / 90.0611	166.1280
Carbamazepine	C <sub>15</sub> H <sub>12</sub> N <sub>2</sub> O	298-46-4	[M+H] <sup>+</sup>	10.84	237.102	237.1020	194.0958	179.0730	165.0700	152.0630	-
Chloramphenicol-D5	C <sub>11</sub> H <sub>7</sub> D <sub>5</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>5</sub>	202480-68-0	[M-H] <sup>-</sup>	6.12	326.036	326.0362	157.0664	34.9695	45.9936	261.8461	326.0340
Chlorphenol (sum)	C <sub>6</sub> H <sub>5</sub> ClO	95-57-8	[M-H] <sup>-</sup>	7.75	126.996	126.9955	34.9694	91.0184	-	-	-
Ciprofloxacin	C <sub>17</sub> H <sub>18</sub> FN <sub>3</sub> O <sub>3</sub>	85721-33-1	[M+H] <sup>+</sup>	5.70	332.140	332.1400	314.1295	231.0570	332.1387	288.1498	98.9847
Climbazole	C <sub>15</sub> H <sub>17</sub> ClN <sub>2</sub> O <sub>2</sub>	38083-17-9	[M+H] <sup>+</sup>	10.14	293.105	293.1048	69.0699	197.0726	141.0104	225.0671	-
Clothianidin	C <sub>6</sub> H <sub>8</sub> ClN <sub>5</sub> O <sub>5</sub> S	2180880-92-5	[M+H] <sup>+</sup>	6.15	250.016	250.0161	131.9673	113.0168	169.0547	110.0722	70.9959
Clothianidin-D3	C <sub>6</sub> H <sub>5</sub> D <sub>3</sub> ClN <sub>5</sub> O <sub>5</sub> S	1262776-24-8	[M+H] <sup>+</sup>	6.12	253.035	253.0349	126.0108	90.0340	253.0316	-	-
Cyproconazol	C <sub>15</sub> H <sub>15</sub> ClN <sub>3</sub> O	94361-06-5	[M+H] <sup>+</sup>	13.73	292.121	292.1209	70.0399	125.0152	138.9956	-	-
Diclofenac-D4	C <sub>14</sub> H <sub>7</sub> D <sub>4</sub> NO <sub>2</sub> Cl <sub>2</sub>	153466-65-0	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	14.11	298.035	298.0343	218.0674 / 34.9691	254.0444	-	-	-
Diethylene glycol dimethyl ether/ Diglyme	C <sub>8</sub> H <sub>14</sub> O <sub>3</sub>	111-96-6	[M+H] <sup>+</sup>	4.62	135.102	135.1015	59.0495	84.9597	116.9289	-	-
Diphenyl sulfone	C <sub>12</sub> H <sub>10</sub> O <sub>2</sub> S	127-63-9	[M+NH <sub>4</sub> ] <sup>+</sup>	10.15	236.074	236.0742	77.0390	219.0493	141.0004	-	-
Ditolyethersulfonic acid	C <sub>14</sub> H <sub>14</sub> O <sub>7</sub> S <sub>2</sub>	-	[M-H] <sup>-</sup>	3.97	357.011	357.0110	277.0550	185.9991	122.0378	79.9590	-
Diuron	C <sub>9</sub> H <sub>10</sub> Cl <sub>2</sub> N <sub>2</sub> O	330-54-1	[M+H] <sup>+</sup>	11.98	233.024	233.0241	72.0444	233.0238	159.9713	132.9614	187.9656
Diuron-D6	C <sub>9</sub> H <sub>4</sub> D <sub>6</sub> Cl <sub>2</sub> N <sub>2</sub> O	1007536-67-5	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	11.95 / 11.93	237.047	239.0619	78.0823 / 185.9522	52.1028 / 34.9693	239.0608	- / 41.9985	- / 149.9749
Fluopyram	C <sub>16</sub> H <sub>11</sub> ClF <sub>3</sub> N <sub>2</sub> O	658066-35-4	[M+H] <sup>+</sup>	13.82	397.054	397.0533	173.0215	208.0140	397.0544	1445.0260	130.0300
Imidacloprid	C <sub>9</sub> H <sub>10</sub> ClN <sub>5</sub> O <sub>2</sub>	138261-41-3	[M+H] <sup>+</sup>	6.70	256.060	256.0593	175.0978	209.0589	84.0541	128.0266	-
Ircadine	C <sub>12</sub> H <sub>23</sub> NO <sub>3</sub>	119515-39-7	[M+H] <sup>+</sup>	12.12	230.175	230.1749	130.1223	95.0856	112.1122	84.0808	156.1025
Lactofen	C <sub>19</sub> H <sub>15</sub> ClF <sub>3</sub> NO <sub>7</sub>	77501-63-4	[M+NH <sub>4</sub> ] <sup>+</sup>	16.53	479.080	479.0823	344.9975	222.9776	-	-	-
Maleic acid	C <sub>4</sub> H <sub>4</sub> O <sub>4</sub>	110-16-7	[M-H] <sup>-</sup>	1.10	115.004	115.0035	-	-	-	-	-
Mecoprop-D3	C <sub>10</sub> D <sub>3</sub> H <sub>8</sub> ClO <sub>3</sub>	352431-15-3	[M-H] <sup>-</sup>	12.23	216.051	216.0512	216.0328	34.9694	197.0487	-	-

Table-A 3-1 continued.

Mesotrione-D4	C14H9D4NO7S	-	[M+H] <sup>+</sup> /[M-H] <sup>-</sup>	8.52 / 8.53	342.0594	344.0737 / 342.0594	-	-	-	-	-
Metsulfuron-methyl-D3	C14H12D3N5O6S	-	[M+H] <sup>+</sup> /[M-H] <sup>-</sup>	10.18 / 10.17	383.086	385.1003 / 383.0859	170.0759 / 142.0813	135.0436	199.0081	209.9831	-
DEF 100*	C12H8N2OS	-	[M+H] <sup>+</sup>	15.65	229.043	229.0429	110.9904	82.9951	-	-	-
Morpholine	C4H9NO	110-91-8	[M+H] <sup>+</sup>	0.64	88.076	88.0754	45.0333	70.0654	42.0338	55.0415	-
Moxifloxacin	C21H24FN3O4	151096-09-2	[M+H] <sup>+</sup>	7.70	402.182	402.1820	402.1824	384.1664	403.1892	364.1619	261.1060
N-(M-Tolyl)-diethanolamine	C11H17NO2	91-99-6	[M+H] <sup>+</sup>	3.72	169.133	169.1327	120.0807	134.0961	91.0542	45.0334	-
N,N-Diethyl-3-methylbenzamide/ DEET	C12H17NO	134-62-3	[M+H] <sup>+</sup>	11.67	192.138	192.1379	91.0538	119.0486	44.0130	-	-
N,N-Diethyl-3-methyl-D3-benzamide-D4/ DEET-D7	C12H10D7NO	1219799-37-7	[M+H] <sup>+</sup>	11.60	199.182	199.1822	126.0941	98.0985	58.9428	199.1831	44.0122
N,N-Dimethyl-N'-phenylsulfamid/ DMSA	C8H12N2O2S	4710-17-2	[M+H] <sup>+</sup>	7.42	201.069	201.0693	92.0496	65.0389	137.1071	-	-
N-Allylthiourea	C4H8N2S	109-57-9	[M+H] <sup>+</sup>	2.08	117.048	117.0481	40.9699	84.9589	58.0662	-	-
N-Nitrosomorpholine	C4H8N2O2	59-89-2	[M+H] <sup>+</sup>	2.51	117.066	117.0657	45.0339	86.0599	57.0569	117.0661	91.0542
N-Phenylurea	C7H8N2O	64-10-8	[M+H] <sup>+</sup>	3.84	137.071	137.0708	77.0389	104.0496	51.0228	119.0604	137.0714
Phthalic acid	C8H6O4	88-99-3	[M-H] <sup>-</sup>	3.67	165.019	165.0193	77.0394	121.0296	96.9602	164.8977	147.8929
Phthalic anhydride	C8H4O3	85-44-9	[M+H] <sup>+</sup>	15.98	149.023	149.0232	65.0388	39.0229	121.0293	84.9602	149.0238
Phthalimide	C8H5NO2	85-41-6	[M+H] <sup>+</sup>	5.43	148.039	148.0392	130.0336	102.0341	75.0235	120.0446	51.0239
Picolinafen	C19H12F4N2O2	137641-05-5	[M+H] <sup>+</sup>	17.02	377.091	148.0392	102.0339	30.0331	75.0237	148.0400	-
Prothicoconazole	C14H15Cl2N3OS	178928-70-6	[M+H] <sup>+</sup>	14.74	344.039	344.0387	98.9843	125.0152	189.0471	326.0312	-
Py-ethylamine	C8H8ClF3N2	-	[M+H] <sup>+</sup>	4.51	225.04	225.0400	196.0136	208.0142	172.0374	160.0371	30.0339
Salicylic acid	C7H6O3	69-72-7	[M-H] <sup>-</sup>	6.29	137.024	137.0243	93.0348	65.0399	-	-	-
Salicylic acid-D4	C7H2D4O3	78646-17-0	[M-H] <sup>-</sup>	6.22	141.05	141.0494	97.0595	69.0648	44.9981	-	-
Spiroxamine	C18H35NO2	118134-30-8	[M+H] <sup>+</sup>	12.00	298.274	298.2739	144.1379	100.1120	126.1278	-	-
Sulcotrione-D4	C14D4H9ClO5S	-	[M+H] <sup>+</sup> /[M-H] <sup>-</sup>	9.71 / 9.70	331.035	333.0493 / 331.0348	-	-	-	-	-
Tebuconazole	C16H22ClN3O	107534-96-3	[M+H] <sup>+</sup>	14.94	308.152	308.1524	70.0397	125.0155	151.0313	165.0467	-
Tebuconazole-D6	C16H16D6ClN3O	-	[M+H] <sup>+</sup>	14.91	314.19	314.1903	72.0537	279.2316	95.0853	314.1824	-
Tetraethyleneglycoldimethylether/ Triglyme	C10H22O5	143-24-8	[M+H] <sup>+</sup>	6.82	223.154	223.1540	59.0489	103.0755	207.0331	191.0017	-
Thiacloprid	C10H9ClN4S	111988-49-9	[M+H] <sup>+</sup>	8.63	253.031	253.0305	126.0108	90.0340	-	-	-
Thiourea	CH4N2S	62-56-6	[M+H] <sup>+</sup>	0.71	77.017	77.0168	59.9901	43.0291	56.9638	-	-
Triethylamine	C6H15N	121-44-8	[M+H] <sup>+</sup>	1.88	102.128	102.1275	58.0649	102.1280	74.0968	30.0339	44.0496
Triethyleneglycoldimethylether/ Triglyme	C8H18O4	112-49-2	[M+H] <sup>+</sup>	5.80	179.128	179.1277	59.0493	103.0759	77.0399	133.0658	-
Trimethylcarboxylic acid	C12H13NO2	-	[M+H] <sup>+</sup>	8.80	204.102	204.1015	188.0706	174.0552	44.0810	160.1124	143.0728

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-2 – Table of concentrations in  $\mu\text{g L}^{-1}$  of spiked compounds for the development of the data evaluation method.

<b>Matrices</b>	<b>Concentration of targets in <math>\mu\text{g L}^{-1}</math></b>	<b>Concentration of ISTD in <math>\mu\text{g L}^{-1}</math></b>
H <sub>2</sub> O	5.00, 50.0	
Eluent A/B	5.00, 10.0	
Matrices of sample stations:	1	10.0, 25.0
	2	25.0, 50.0
		5.00, 10.0,
	3	
	4	25.0, 50.0
	5.00, 25.0	Bezafibrate-D6 (25.0), Diclofenac-D4 (25.0), Mesotrione-D4 (25.0), Metsulfuron-methyl-D3 (25.0), Sulcotrione-D4 (0.250), Diuron-D6 (25.0), Clothianidin-D3 (25.0), Tebuconazol-D6 (1.25), DEET-D7 (1.25), Azoxystrobin-D4 (0.0250), Salicylsäure-D4 (25.0), Mecoprop-D3 (25.0), Chloramphenicol-D5 (25.0)
5	5.00, 10.0,	
	25.0, 50.0	

### 3.6.1.2 Non-target screening method

Table-A 3-3 – Parameter settings of 'non-target analysis' determination with SCEIX OS.

<b>'Non-target analysis' parameter</b>	<b>Settings</b>
<b>a)</b>	
Minimum Peak Width	6 points
Minimum Peak Height	100
S/N Integration Threshold	10
<b>b)</b>	
XIC width	0.05 Da
Gaussian Smooth width	1 point
Noise Percentage	90%
Baseline Subtract Window	0.3 min
Peak Splitting	2 points
<b>c)</b>	
RT Half Window	9 sec
<b>d)</b>	
Precursor mass tolerance	$\pm 2$ mDa
Fragment mass tolerance	$\pm 50$ mDa
Collision energy	$\pm 5$ eV
Use polarity	Yes
Use collision energy spread	Yes
Maximum number of Hits	5
Intensity threshold	0.05
Minimal purity required for matching	20%
Intensity factor (In the library a spectrum is searched with the same relative intensity as the unknown spectrum $\pm$ the intensity factor)	5
<b>e)</b>	
Non-Target Peaks	1 – 26 min
Only signals with the specified spectral intensity will be put on the component list and considered for library search and formula finder	Exhaustive
Area Ratio Threshold (Unknown/Control):	1000
Group Peaks by adduct or charge	Yes

Table-A 3-4 – Table of settings used for data-independent acquisition (vDIA).

<b>vDIA parameters</b>	<b>Settings</b>
Target number of windows	15
Lower <i>m/z</i> limit	200
Upper <i>m/z</i> limit	960
Window overlap (Da)	1 Da
Minimum window width	5 Da
Collision energy spread (CES)	15

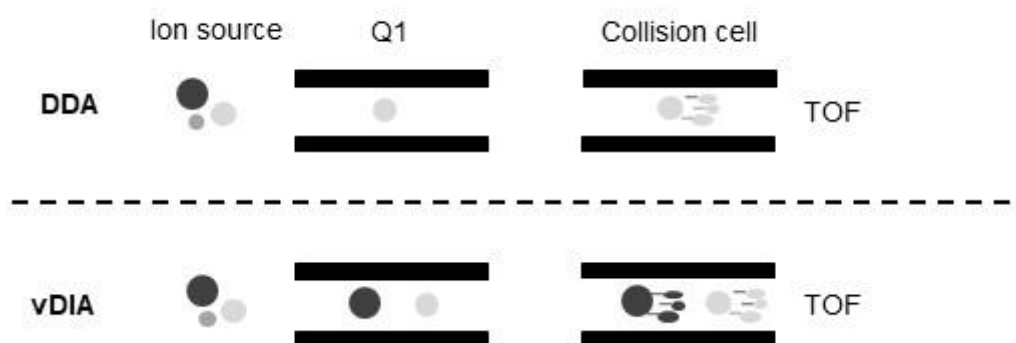


Figure-A 3-1 – Illustration of different fragmentation methods (DDA  $\triangleq$  data-dependent acquisition, vDIA  $\triangleq$  variable data-independent acquisition) used in the method development.

### 3.6.1.3 ISTD stability over Batch

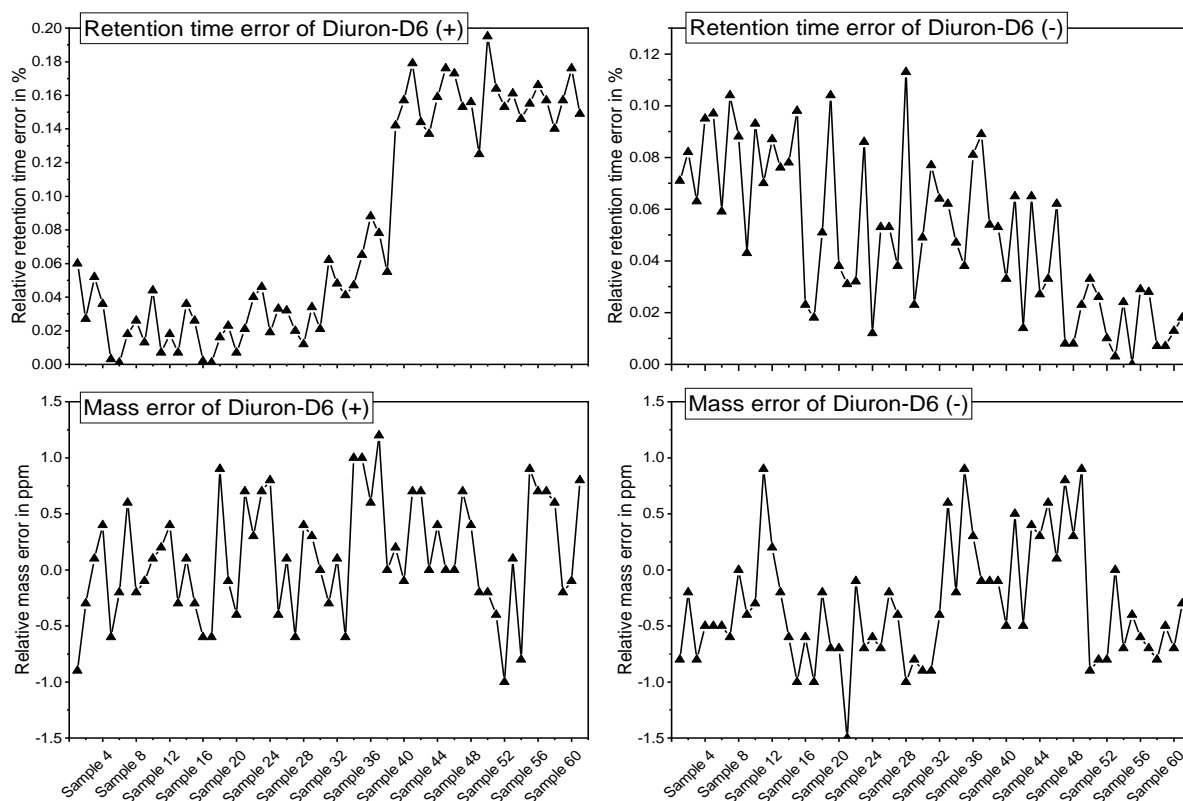


Figure-A 3-2 – Illustration of stability of ISTD (exemplarily shown for Diuron-D6) over one batch for quality control.

To assess the robustness of the method, all ISTD of Table-A 3-2 were tested for stability throughout the analysis batch. From these experiments, all retention times were found stable. For example, the relative standard deviation calculated for the retention time of the internal standard diuron-D6 in positive ionisation mode is < 1.8%. The mass accuracy also showed excellent stability. The accurate mass measurement of the internal standard diuron-D6 ( $m/z$  239.06196) was evaluated over the batch and the deviation of the accurate mass ranged from  $-0.7$  to  $1.0$  ppm ( $0.16$  to  $0.24$  mDa) throughout seven days of validation (see Table-A 3-5).

Table-A 3-5 – Method validation results of internal standards in positive (+) and negative (-) ionisation mode measured over seven days.

	<b>Component Name</b>	<b>The relative standard deviation of retention time in %</b>	<b>Maximum mass error tolerance in ppm</b>	<b>Minimum mass error tolerance in ppm</b>	<b>Maximum mass error tolerance in mDa</b>	<b>Minimum mass error tolerance in mDa</b>
(+)	Bezafibrate-D6	1.37	0.00	-1.00	0.01	-0.41
	Chloramphenicol-D5	2.83	1.00	-0.10	0.37	-0.05
	Diclofenac-D4	3.02	0.00	-1.90	0.01	-0.37
	Diuron-D6	1.81	1.40	-0.20	0.43	-0.05
	Mecoprop-D3	3.08	1.00	-0.70	0.24	-0.16
	Mesotrione-D4	1.99	1.40	-0.80	0.50	-0.27
	Metsulfuron-methyl-D3	3.14	1.20	-0.10	0.48	-0.04
	Salicylic Acid-D4	1.89	0.60	-0.90	0.22	-0.29
	Sulcotrione-D4	3.21	1.60	-0.60	0.49	-0.18
(-)	Azoxystrobin-D4	0.92	2.50	-1.90	0.93	-0.69
	Bezafibrate-D6	0.82	1.00	-2.10	0.32	-0.67
	Clothianidin-D3	0.56	10.00	-6.00	2.22	-1.32
	DEET-D7	0.69	0.80	-1.00	0.23	-0.29
	Diclofenac-D4	1.02	1.00	-1.20	0.23	-0.28
	Diuron-D6	0.75	1.00	-1.20	0.22	-0.25
	Mesotrione-D4	0.68	1.00	-1.70	0.34	-0.58
	Metsulfuron-methyl-D3	0.70	1.00	-1.00	0.38	-0.38
	Sulcotrione-D4	0.65	1.30	-0.80	0.18	-0.11
	Tebuconazole-D6	1.10	0.90	-1.20	0.29	-0.40

### 3.6.2 Supplement for Chapter Results

#### 3.6.2.1 Centrifugation

Table-A 3-6 – Results of the experiments of centrifugation of 17 inflow samples of the WWTP (SP 3), 12 pre-treated samples (SP 4) and 16 samples of outflow of the WWTP (SP 5). The relative standard deviation (RSD) is the mean of nine measurements and represents the deviation of centrifugation experiments to initial measurements without centrifugation.

<b>Target compounds</b>	<b>RSD in % of samples from</b>		
	<b>SP 3</b>	<b>SP 4</b>	<b>SP 5</b>
1-(3'-Sulfophenyl)-3-methyl-5-aminopyrazole	78.1	76.1	11.2
1,5-Diaminonaphthalene	10.2	74.0	-
1,7-Naphthalenedisulfonic acid	7.67	52.0	40.3
2-(n-Ethylanilino)ethanol	3.09	0.940	6.19
2-(Trifluoromethyl)benzenesulfonamide/ TBSA	3.53	2.65	1.82
2-(Trifluoromethyl)benzamide	9.73	12.2	15.6
2,2'-(4-Methylphenylimino)diethanol	0.940	3.29	4.39
2,4,6-Triamino-1,3,5-triazine/ Melamine	12.1	5.46	7.47
2,4,6-Trimethoxy-1,3,5-triazine	6.25	2.77	1.92
2-Amino-4-methoxy-6-trifluoromethyl-1,3,5-triazine/ AMTT	5.73	10.5	12.2
2-Chlorobenzoic acid	57.7	-	57.7
2-Naphthalenesulfonic acid	3.63	3.06	1.03
3-Hydroxy-2-nitrobenzoic acid	9.69	5.23	10.6
4-Aminobenzoic acid	16.9	4.82	4.65
4-Chlor-2-cyano-5-(4-methylphenyl)imidazole/ CCIM	4.10	1.98	5.65
4-Chlorobenzoic acid	5.50	16.9	67.0
4-Toluolsulfonic acid	52.7	1.94	6.09
6-Chloropyridine-2-carboxylic acid	12.5	0.97	3.97
Acifluofen	19.2	3.66	10.1
Azoxystrobin	5.26	3.29	5.54
ABC 700	3.65	12.0	10.5

Table-A 3-6 continued.

Benzamideoxime	2.71	203	2.12
Benzenesulfonic acid	5.58	3.08	4.34
Carbamazepine	34.5	3.69	5.50
Chlorphenol (sum)	157	17.4	30.5
Ciprofloxacin	11.6	6.03	2.91
Climbazole	18.2	8.22	3.27
Clothianidin	8.84	6.94	9.49
Cyproconazole	5.10	6.15	8.66
Diethylene glycol dimethyl ether/ Diglyme	1.38	14.01	4.66
Diphenyl sulfone	5.07	8.03	1.71
Ditolyetherdisulfonic acid	14.8	11.4	5.34
Fluopyram	5.74	6.45	8.93
Imidacloprid	12.5	4.97	14.2
Lactofen	51.4	71.9	44.8
Metanilic acid	57.7	-	4.70
DEF 100	29.8	5.29	4.82
Morpholine	8.49	8.28	16.3
Moxifloxacin	16.2	4.36	4.75
N-(M-Tolyl)-diethanolamine	5.64	2.94	3.61
N,N-Diethyl-3-methylbenzamide/ DEET	0.592	3.22	2.70
N,N-Dimethyl-N'-phenylsulfamid/ DMSA	7.67	4.71	7.17
N-Nitrosomorpholine	11.6	15.7	115
Phthalic acid	0.554	57.7	3.02
Phthalimide	7.40	-	-
Picolinafen	51.1	37.7	25.3
Prothioconazole	5.74	1.45	8.31
Py-ethylamine	51.6	2.63	6.36
Salicylic acid	174	4.01	8.45
Saltidine	4.81	2.01	5.17
Spiroxamine	42.5	15.2	7.54
Tebuconazole	4.50	3.18	4.13
Tetraethyleneglycoldimethylether/ Tetraglyme	1.69	2.04	4.88
Thiacloprid	9.55	10.9	8.22
Triethyleneglycoldimethylether/ Triglyme	3.31	5.06	3.01

### 3.6.2.2 Sample Dilution

In Figure-A 3-3 a) and Figure-A 3-3 b) the comparison of the count conversion factor (CCF) of the dynamic ion transmission control (ITC), of in the first case a pre-treated wastewater sample (SP 4) and in the second case a diluted WWTP influent sample (SP 2) with respectively the undiluted samples, are shown. The ITC is an application in TOF-MS which modulates the ion current through the instrument scan by scan protecting the detector against saturation. A count conversion factor of 100% indicates the instrument is not getting overloaded with the signal. In contrast, a modulation down to 2% arises for an overloaded detector. The CCF is the time (in seconds) ions are being transmitted during each cycle. For example, if there is a TOF-MS experiment with a 250 ms accumulation time and ITC was set to 10% for a cycle, then the CCF will be  $0.10 * 0.25 \text{ s} = 0.025 \text{ s}$ . Conversely, if there is a CCF value of 0.0052s and the accumulation time is 150 ms, then the ITC is calculated as  $0.0052 \text{ s} / 0.15 \text{ s} = 0.0347 = 3.47\%$ . In the presented case, the accumulation time was 0.1 s. In Figure-A 3-3 a) the ITC is modulated to about 10%. In comparison, for the influent sample, the ITC is set to 2%, which shows distinct overloading. The detector was saturated, demanding less sample amount should be loaded. The pre-treated wastewater sample (SP 4) also showed more modulations for the sample without dilution than with dilution but in a much more decisive way compared to



the WWTP influent. Consequently, all the samples of SP 1 to SP 5 were diluted before analysis.

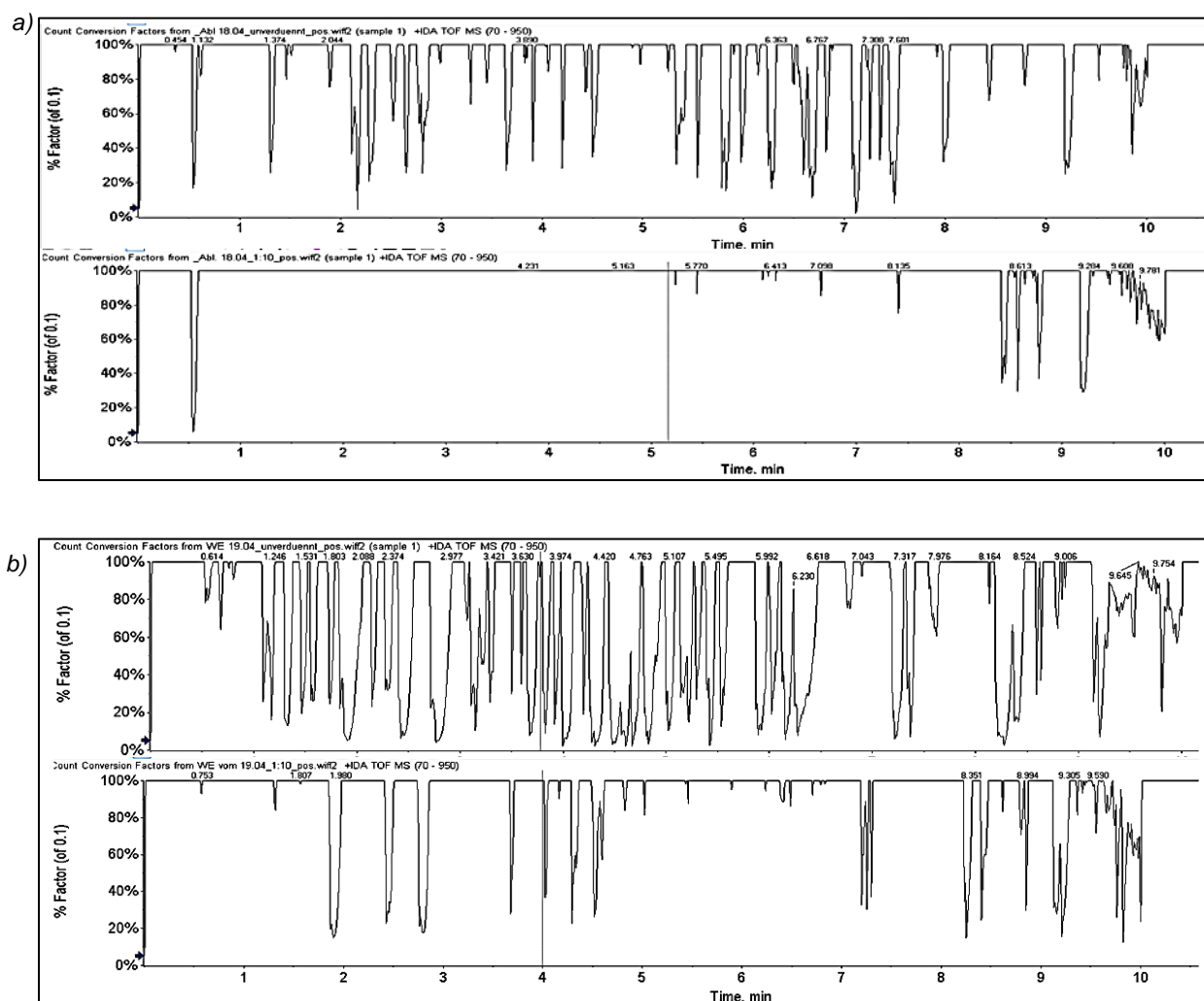


Figure-A 3-3 – Comparison of the CCF from a) an exemplary undiluted pre-treated wastewater sample (SP 4, first chromatogram) and the same sample with dilution (1 to 10 with Eluent A/B (V/V), second chromatogram) and b) an exemplary WWTP influent sample (SP 2) without any dilution (first chromatogram) and diluted 1 to 10 (second chromatogram). The y-axis presents the ITC calculated by the CCF divided by the accumulation time (in the given case: 0.1 s).

3.6.2.3 Stability Analysis during Sample Storage

Table-A 3-7 – Results of stability analysis during sample storage. The storage conditions of freezing and cooling were tests for samples of sampling sites 3 to 5. In the table, the recovery in % is presented.

Target Compounds	Cooling at 8 °C			Freezing at – 22 °C		
	SP 3	SP 4	SP 5	SP 3	SP 4	SP 5
Iminopyrazole Acid-3	0-10%	0-10%	0-10%	80-60%	80-60%	80-120%
1.5-Diaminonaphthalene	80-60%	10-40%	10-40%	80-120%	80-120%	80-120%
TBSA	80-60%	80-60%	40-50%	80-120%	80-120%	80-120%
2-(Trifluoromethyl)benzamide	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Melamine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
2.4.6-Trimethoxy-1.3.5-triazine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
AMTT	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
2-Chlorobenzoic Acid	0-10%	0-10%	10-40%	80-60%	80-120%	80-120%
3-Hydroxy-2-nitrobenzoic Acid	10-40%	10-40%	40-50%	80-120%	80-120%	80-120%
4-Aminobenzoic Acid	0-10%	0-10%	0-10%	80-60%	80-60%	80-60%
CCIM	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
4-Chlorobenzoic Acid	80-60%	40-50%	40-50%	80-120%	80-120%	80-120%
4-Toluolsulfonic Acid	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
6-Chloropicolinic Acid	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Acifluorfen	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Azoxystrobin	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
ABC 700	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Benzamide oxime	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Benzenesulfonic Acid	0-10%	0-10%	0-10%	80-60%	80-60%	80-120%
Carbamazepine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Chlorphenol (sum)	10-40%	10-40%	10-40%	80-120%	80-120%	80-120%
Ciprofloxacin	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Climbazole	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Clothianidin	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Cyproconazole	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Diglyme	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Diphenyl sulfone	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Ditolyetherdisulfonic Acid	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Fluopyram	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Imidacloprid	80-60%	80-60%	40-50%	80-120%	80-120%	80-120%
Lactofen	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Metanilic Acid	10-40%	10-40%	10-40%	80-120%	80-120%	80-120%
DEF 100	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Morpholine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Moxifloxacin	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
DEET	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
DMSA	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
N.N-Bis-(2-hydroxyethyl)-3-toluidine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
N.N-Dihydroxyethyl-4-toluidine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Naphthalene-1.7-disulfonic Acid	40-50%	40-50%	40-50%	10-40%	80-60%	80-60%
Naphthalene-2-sulfonic Acid	80-60%	80-60%	40-50%	80-120%	80-120%	80-120%
N-Ethyl-N-(2-hydroxyethyl)-aniline	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
N-Nitrosomorpholine	10-40%	10-40%	10-40%	80-120%	80-120%	80-120%
Phthalic Acid	80-60%	80-60%	40-50%	80-120%	80-120%	80-120%
Phthalimide	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Picolinafen	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Prothioconazole	10-40%	10-40%	10-40%	80-120%	80-120%	80-120%
Py-ethylamine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Salicylic Acid	80-60%	40-50%	40-50%	80-120%	80-120%	80-120%
Saltidine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Spiroxamine	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Tebuconazole	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Tetraglyme	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Thiacloprid	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%
Triglyme	40-50%	40-50%	40-50%	80-120%	80-120%	80-120%
Trimethylcarboxylic Acid	80-60%	80-60%	80-60%	80-120%	80-120%	80-120%

### 3.6.2.4 Long-term Storage Conditions

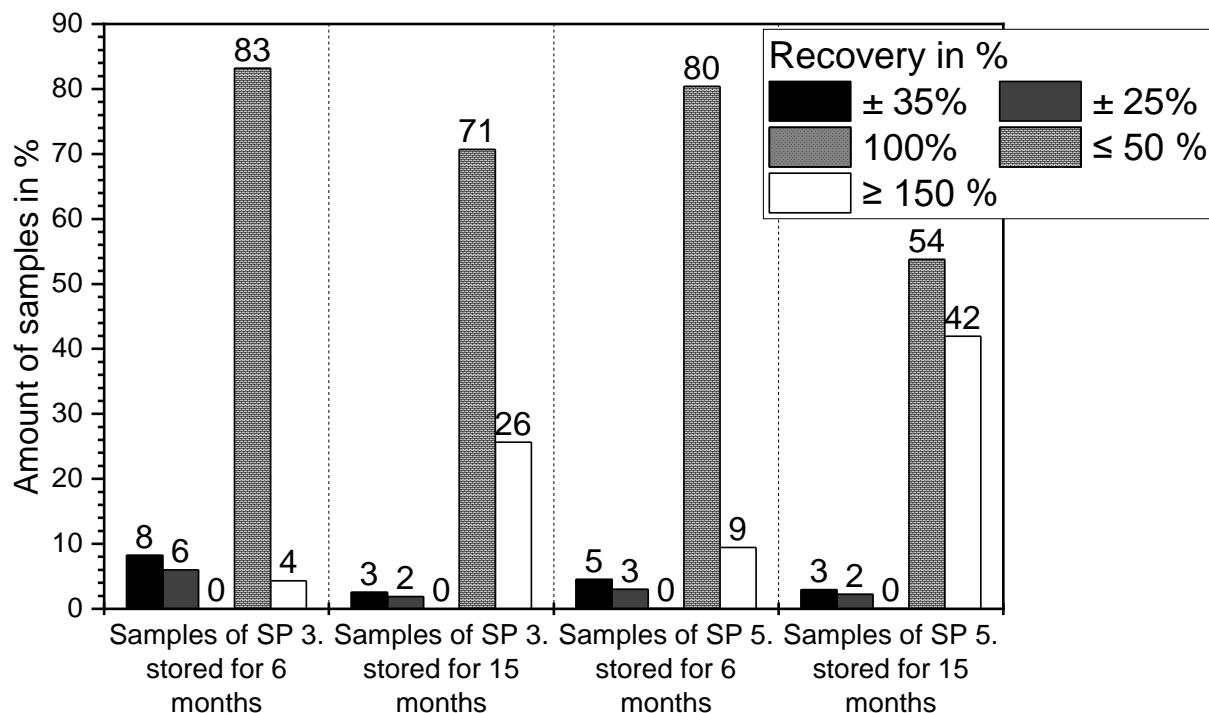


Figure-A 3-4– Recovery of the WWTP inlet (SP 3) and the WWTP outflow (SP 5) of authentic samples from January 2018 (2018-01-01 to 2018-01-31) after six months and 15 months. The recovery is based on all features (target approach) which were extracted in January 2018 in the suspected-target evaluation. The data label of bars indicates the defined sample amount covering the recovery in % (cf. y-axis). The bars illustrate the recovery in %. The results represent the mean values of 31 samples for each matrix.

### 3.6.2.5 LC-Column Tests

During the method development, tests were carried out to obtain the best combination of stationary phase (column) and mobile phase. Four columns were tested in combination with the following eluents as a mobile phase:

Table-A 3-8 – Tested mobile phases during method development.

Mobile Phase	Eluent A	Eluent B
1	0.01% formic acid + 5 mM ammonium formate in Milli-Q® water	0.01% formic acid + 5 mM ammonium formate in methanol
2	0.1% formic acid in Milli-Q® water	0.1% formic acid in methanol
3	0.1% formic acid in Milli-Q® water	0.1% formic acid in acetonitrile
4	0.1% formic acid in Milli-Q® water	Acetonitrile
5	10 mM ammonium formate in Milli-Q® water	0.1% formic acid in methanol
6	0.1% formic acid + 5 mM ammonium formate in Milli-Q® water	0.1% formic acid in methanol

Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-9 – Recovery of target analytes during optimisation of LC-column. In the table, the mean areas of three injections of 10 µg L<sup>-1</sup> standard mixture with corresponding standard deviation (SD) are shown. The results based on mobile phase 2 (see Table-A 3-8).

Target compounds	Kinetex		Synergi		Raptor Biphenyl		Ultra Aqueous	
	Mean area in cps	SD	Mean area in cps	SD	Mean area in cps	SD	Mean area in cps	SD
Iminopyrazole Acid-3	147	0.20	115	0.13	193	0.09	2253	0.04
1,5-Diaminonaphthalene	12486	0.12	1784	0.15	11557	0.11	8721	0.05
2-(n-Ethylanilino)ethanol	19919	0.04	29970	0.02	44349	0.09	49946	0.09
TBSA	2451	0.10	2035	0.03	2456	0.10	2185	0.01
2-(Trifluoromethyl)benzamide	6859	0.03	9856	0.13	6913	0.09	8695	0.02
2,2'-(4-Methylphenylimino)diethanol	49893	0.10	47327	0.07	476	0.10	49410	0.00
Melamine	53	0.09	17	0.09	19109	0.00	18654	0.05
AMTT	27296	0.12	16881	0.15	23036	0.06	29243	0.03
6-Chloropyridine-2-carboxylic acid	1513	0.19	573	0.02	2180	0.05	2277	0.07
Aciflurofen	1936	0.19	754	0.01	465	0.11	1154	0.07
Azoxystrobin	67691	0.00	20872	0.11	25989	0.05	40488	0.10
ABC 700	5457	0.09	7649	0.15	6006	0.09	7800	0.03
Benzamideoxime	8942	0.00	1834	0.01	12401	0.00	17970	0.00
Benzoic acid	464	0.06	171	0.09	668	0.04	874	0.10
Benzenesulfonic acid	6092	0.19	7738	0.05	6345	0.01	6433	0.10
Carbamazepine	77943	0.18	36565	0.09	55629	0.05	68186	0.03
Ciprofloxacin	1026	0.10	990	0.11	1484	0.08	1324	0.02
Climbazole	61736	0.16	54099	0.09	47120	0.11	48932	0.03
Clothianidin	3949	0.13	2419	0.13	2659	0.09	3872	0.05
Cyproconazole	34907	0.04	8908	0.08	19323	0.03	20647	0.03
Diglyme	679	0.10	0	0.13	680	0.03	1083	0.08
Diphenyl sulfone	2778	0.17	12324	0.10	28266	0.04	26101	0.07
Fluopyram	33680	0.19	15101	0.14	32799	0.12	47917	0.07
Imidacloprid	8503	0.18	8878	0.04	5688	0.06	7175	0.08
Lactofen	2268	0.02	258	0.09	978	0.05	2353	0.08
Metanilic acid	187	0.18	226	0.12	992	0.08	2997	0.01
DEF 100	3272	0.13	1194	0.08	4417	0.00	4911	0.09
Morpholine	3979	0.19	3925	0.10	5664	0.07	6179	0.02
Moxifloxacin	1107	0.07	5716	0.04	4588	0.04	4523	0.04
N-(M-Tolyl)-diethanolamine	53171	0.20	53861	0.11	50954	0.10	53238	0.08
DEET	106714	0.00	96529	0.10	107770	0.04	114676	0.08
DMSA	7262	0.20	3889	0.14	6023	0.10	7809	0.09
N-Nitrosomorpholine	142	0.10	370	0.07	1681	0.09	3518	0.01
Phthalic acid	1064	0.10	767	0.00	1235	0.06	1279	0.07
Picolinafen	31250	0.06	4063	0.03	16196	0.07	34988	0.07
Prothioconazole	2831	0.08	743	0.08	3781	0.03	3394	0.06
Py-ethylamine	30332	0.19	13986	0.14	23913	0.09	28454	0.08
Salicylic acid	4193	0.07	4232	0.09	4382	0.12	6344	0.04
Tebuconazole	44405	0.06	15470	0.02	30315	0.02	37462	0.04
Tetraglyme	14293	0.14	5510	0.04	14104	0.10	15855	0.08
Thiacloprid	19270	0.15	13770	0.00	13473	0.11	27619	0.04
Triglyme	5957	0.14	7655	0.15	8100	0.12	8495	0.02

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-10 – Retention (capacity) factor ( $k = \frac{(t_R - t_0)}{t_0}$ ) of target analytes during optimisation of LC-column. In the table, the mean  $k$  of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) are shown. The void time for calculation based on each column on the injection peak (0.9 min). The capacity factor shows the retention of an analyte on the chromatographic column. A high  $k$ -value indicates that the analyte is highly retained and has spent a significant amount of time interacting with the stationary phase. The results based on mobile phase 2 (see Table-A 3-8).

Target compounds	Kinetex		Synergi		Raptor Biphenyl		Ultra Aqueous	
	$k$	SD	$k$	SD	$k$	SD	$k$	SD
Iminopyrazole Acid-3	13.4	0.02	11.3	0.01	6.9	0.04	18.4	0.04
1,5-Diaminonaphthalene	0.8	0.00	1.1	0.05	2.1	0.05	0.8	0.03
2-(n-Ethylanilino)ethanol	2.3	0.06	2.8	0.08	3.8	0.05	2.3	0.04
TBSA	5.1	0.08	5.6	0.08	6.4	0.08	5.1	0.02
2-(Trifluoromethyl)benzamide	5.1	0.08	6.6	0.08	6.1	0.04	5.1	0.01
2,2'-(4-Methylphenylimino)diethanol	3.9	0.07	4.9	0.04	5.0	0.07	3.9	0.03
Melamine	3.9	0.01	4.9	0.07	0.1	0.04	3.9	0.02
AMTT	2.3	0.01	7.8	0.03	6.1	0.05	6.3	0.04
6-Chloropyridine-2-carboxylic acid	4.3	0.02	5.6	0.04	6.1	0.06	4.3	0.02
Aciflurofen	9.8	0.05	10.8	0.06	13.1	0.06	12.8	0.01
Azoxystrobin	10.1	0.02	14.3	0.07	15.9	0.07	13.1	0.02
ABC 700	3.4	0.01	4.4	0.07	4.3	0.02	3.4	0.04
Benzamideoxime	0.8	0.04	0.9	0.07	1.3	0.06	0.8	0.03
Benzoic acid	1.0	0.05	2.0	0.03	1.3	0.03	1.0	0.02
Benzenesulfonic acid	3.5	0.04	4.1	0.05	4.4	0.04	3.5	0.03
Carbamazepine	8.8	0.07	10.0	0.05	12.5	0.00	10.8	0.03
Ciprofloxacin	5.4	0.03	5.5	0.09	6.9	0.07	5.4	0.04
Climbazole	9.8	0.01	11.4	0.09	11.4	0.02	10.8	0.03
Clothianidin	5.6	0.01	6.8	0.09	6.8	0.07	5.6	0.04
Cyproconazole	13.9	0.03	11.3	0.08	14.0	0.06	13.9	0.00
Diglyme	3.9	0.08	3.1	0.02	4.9	0.06	3.9	0.01
Diphenyl sulfone	10.5	0.05	11.9	0.07	12.9	0.01	10.5	0.03
Fluopyram	14.3	0.08	15.4	0.03	14.1	0.03	14.3	0.00
Imidacloprid	5.9	0.04	6.9	0.03	9.0	0.05	5.9	0.01
Lactofen	17.0	0.03	18.1	0.04	17.5	0.00	17.0	0.05
Metanilic acid	0.0	0.09	0.0	0.01	0.0	0.06	0.0	0.01
DEF 100	15.5	0.00	12.4	0.02	15.9	0.02	15.5	0.00
Morpholine	0.1	0.07	0.0	0.10	0.0	0.07	0.1	0.03
Moxifloxacin	8.4	0.07	7.6	0.10	9.5	0.06	7.4	0.02
N-(M-Tolyl)-diethanolamine	3.3	0.06	4.0	0.04	4.6	0.06	3.3	0.00
DEET	11.8	0.00	13.3	0.07	13.0	0.02	11.8	0.03
DMSA	8.1	0.00	9.5	0.01	9.6	0.01	8.1	0.01
N-Nitrosomorpholine	1.4	0.01	2.3	0.09	2.8	0.07	1.4	0.00
Phthalic acid	3.5	0.06	4.3	0.06	4.4	0.04	3.5	0.03
Picolinafen	13.1	0.06	16.3	0.03	17.0	0.04	17.1	0.04
Prothioconazole	12.3	0.06	14.5	0.10	15.4	0.02	15.3	0.01
Py-ethylamine	5.4	0.00	5.9	0.06	5.5	0.05	5.4	0.02
Salicylic acid	6.4	0.08	7.1	0.08	6.5	0.04	6.4	0.03
Tebuconazole	15.1	0.01	13.4	0.05	14.9	0.08	15.1	0.05
Tetraglyme	6.1	0.07	7.4	0.06	8.0	0.06	6.1	0.04
Thiacloprid	5.6	0.08	9.9	0.01	11.3	0.01	7.6	0.00
Triglyme	5.1	0.05	3.4	0.03	6.6	0.03	5.1	0.03

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-11 – Asymmetric factor (AF) of target analytes during optimisation of LC-column. In the table, the mean AF of three injections of 10 µg L<sup>-1</sup> standard with corresponding standard deviation (SD) is shown. The asymmetry factor shows the distance from the centre line of the peak to the back slope, divided by the distance from the centre line of the peak to the front slope, with all of the measurements made at 10% of the maximum peak height (AF = 1  $\triangleq$  symmetric peak, AF > 1  $\triangleq$  tailing, AF < 1  $\triangleq$  fronting). The results based on mobile phase 2 (see Table-A 3-8).

Target compounds	Kinetex		Synergi		Raptor Biphenyl		Ultra Aqueous	
	AF	SD	AF	SD	AF	SD	AF	SD
Iminopyrazole Acid-3	0.77	0.01	4.57	0.03	8.26	0.11	0.83	0.09
1,5-Diaminonaphthalene	1.93	0.05	0.48	0.02	1.54	0.03	1.23	0.06
2-(n-Ethylanilino)ethanol	2.5	0.01	2.28	0.01	1.22	0.04	1.23	0.08
TBSA	1.11	0.14	1.08	0.03	1.33	0.06	1.21	0.05
2-(Trifluoromethyl)benzamide	1.04	0.02	2.96	0.02	1.09	0.08	0.87	0.05
2,2'-(4-Methylphenylimino)diethanol	1.53	0.12	3.29	0.05	1.11	0.04	1.05	0.02
Melamine	1.53	0.04	3.3	0.06	1.59	0.05	1.29	0.07
AMTT	0.93	0.08	0.3	0.09	1.32	0.08	1.02	0.06
6-Chloropyridine-2-carboxylic acid	1.78	0.08	2.73	0.00	1.59	0.07	1.18	0.05
Aciflurofen	0.85	0.11	1.45	0.02	2.17	0.10	0.79	0.05
Azoxystrobin	1.29	0.06	0.63	0.05	1.33	0.06	1.05	0.08
ABC 700	1.39	0.02	2.66	0.02	0.97	0.08	1.01	0.02
Benzamideoxime	2.34	0.15	1.68	0.04	1.25	0.02	1.37	0.06
Benzoic acid	1.7	0.01	1.36	0.05	1.19	0.01	1.25	0.01
Benzenesulfonic acid	0.27	0.14	1.08	0.09	0.62	0.03	1.18	0.04
Carbamazepine	1.34	0.07	0.6	0.03	1.08	0.09	1.13	0.01
Ciprofloxacin	2.95	0.11	2.3	0.02	1.45	0.05	0.88	0.04
Climbazole	0.83	0.04	1.05	0.03	1.08	0.09	1.03	0.06
Clothianidin	1.22	0.00	0.73	0.04	1.56	0.08	0.73	0.03
Cyproconazole	0.26	0.14	0.58	0.02	1.25	0.04	0.91	0.01
Diglyme	1.95	0.09	N/A	0.06	0.9	0.11	1.18	0.07
Diphenyl sulfone	1.59	0.00	4.1	0.04	1.26	0.02	1.24	0.03
Fluopyram	1.14	0.15	2.78	0.04	1.21	0.10	0.98	0.09
Imidacloprid	1.05	0.01	1.53	0.06	1.17	0.06	1.07	0.02
Lactofen	1.45	0.04	0.13	0.07	1.14	0.04	1.03	0.01
Metanilic acid	1.83	0.12	6.14	0.09	2.87	0.08	0.96	0.03
DEF 100	1.19	0.11	0.79	0.01	1.74	0.09	1.33	0.09
Morpholine	0.76	0.00	2.39	0.02	1.61	0.03	1.19	0.02
Moxifloxacin	0.9	0.04	1.35	0.05	1	0.06	1.34	0.02
N-(M-Tolyl)-diethanolamine	1.54	0.11	3.22	0.03	1.12	0.02	1.23	0.00
DEET	1.37	0.14	1.27	0.07	1.17	0.00	1.13	0.03
DMSA	1.44	0.09	3.82	0.02	1.34	0.10	0.77	0.03
N-Nitrosomorpholine	0.92	0.03	3.71	0.09	1.03	0.10	0.88	0.04
Phthalic acid	2.16	0.08	8.34	0.05	2.1	0.04	1.08	0.09
Picolinafen	1.5	0.12	0.6	0.01	1.61	0.01	1.23	0.07
Prothioconazole	0.99	0.03	0.6	0.06	0.81	0.02	1.37	0.06
Py-ethylamine	1.38	0.14	4.93	0.09	1.01	0.05	0.97	0.08
Salicylic acid	1.62	0.11	0.31	0.07	1.76	0.03	1.21	0.03
Tebuconazole	1.58	0.06	1.71	0.06	1.37	0.03	1.31	0.07
Tetraglyme	1.75	0.07	1.27	0.08	1.41	0.05	1.11	0.08
Thiacloprid	1.51	0.06	1.37	0.05	1.28	0.09	1.11	0.02
Triglyme	1.01	0.03	1.34	0.07	1.39	0.06	0.91	0.07

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-12 – Results of full width at half maximum (FWHM) of target analytes during optimisation of LC-column. In the table, the mean FWHM of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) is shown. The FWHM shows the chromatographic peak width, in minutes, of the detected peak measured at half of its apex intensity. The results based on mobile phase 2 (see Table-A 3-8).

Target compounds	Kinetex		Synergi		Raptor Biphenyl		Ultra Aqueous	
	Width at 50%	SD	Width at 50%	SD	Width at 50%	SD	Width at 50%	SD
Iminopyrazole Acid-3	0.01	0.06	0.04	0.06	0.01	0.01	0.01	0.07
1,5-Diaminonaphthalene	0.08	0.01	0.02	0.09	0.05	0.06	0.06	0.07
2-(n-Ethylanilino)ethanol	0.06	0.05	0.03	0.09	0.04	0.02	0.05	0.05
TBSA	0.04	0.08	0.06	0.04	0.04	0.03	0.04	0.00
2-(Trifluoromethyl)benzamide	0.04	0.12	0.07	0.01	0.04	0.01	0.05	0.00
2,2'-(4-Methylphenylimino)diethanol	0.04	0.11	0.07	0.09	0.04	0.01	0.04	0.05
Melamine	0.04	0.09	0.07	0.05	0.04	0.06	0.04	0.05
AMTT	0.04	0.02	0.05	0.08	0.05	0.01	0.05	0.01
6-Chloropyridine-2-carboxylic acid	0.04	0.08	0.03	0.05	0.05	0.06	0.05	0.07
Aciflurofen	0.04	0.03	0.08	0.06	0.03	0.06	0.05	0.02
Azoxystrobin	0.04	0.02	0.05	0.06	0.05	0.02	0.05	0.06
ABC 700	0.04	0.11	0.06	0.05	0.04	0.04	0.05	0.01
Benzamideoxime	0.07	0.10	0.1	0.00	0.07	0.03	0.07	0.02
Benzoic acid	0.06	0.05	0.1	0.04	0.06	0.04	0.06	0.08
Benzenesulfonic acid	0.04	0.12	0.01	0.04	0.03	0.06	0.05	0.03
Carbamazepine	0.04	0.10	0.04	0.08	0.05	0.01	0.05	0.00
Ciprofloxacin	0.09	0.02	0.03	0.08	0.07	0.01	0.07	0.00
Climbazole	0.05	0.11	0.06	0.03	0.05	0.07	0.05	0.06
Clothianidin	0.03	0.05	0.04	0.10	0.04	0.09	0.05	0.04
Cyproconazole	0.17	0.11	0.2	0.04	0.04	0.03	0.05	0.08
Diglyme	0.04	0.01	N/A	0.09	0.04	0.06	0.05	0.01
Diphenyl sulfone	0.04	0.05	0.03	0.09	0.05	0.04	0.05	0.04
Fluopyram	0.04	0.09	0.07	0.03	0.04	0.09	0.05	0.00
Imidacloprid	0.04	0.12	0.05	0.03	0.04	0.07	0.05	0.04
Lactofen	0.03	0.07	0.04	0.04	0.03	0.08	0.04	0.00
Metanilic acid	0.04	0.09	0.04	0.04	0.04	0.05	0.03	0.08
DEF 100	0.03	0.12	0.02	0.04	0.05	0.01	0.05	0.08
Morpholine	0.03	0.00	0.05	0.07	0.03	0.10	0.03	0.05
Moxifloxacin	0.07	0.06	0.06	0.02	0.05	0.10	0.06	0.01
N-(M-Tolyl)-diethanolamine	0.05	0.05	0.06	0.03	0.04	0.05	0.05	0.04
DEET	0.04	0.00	0.05	0.05	0.04	0.07	0.05	0.08
DMSA	0.03	0.10	0.04	0.01	0.05	0.09	0.05	0.07
N-Nitrosomorpholine	0.01	0.01	0.01	0.03	0.02	0.05	0.03	0.04
Phthalic acid	0.05	0.07	0.07	0.06	0.05	0.08	0.05	0.06
Picolinafen	0.04	0.09	0.04	0.08	0.04	0.02	0.04	0.04
Prothioconazole	0.05	0.10	0.02	0.08	0.04	0.06	0.05	0.04
Py-ethylamine	0.05	0.12	0.09	0.03	0.05	0.09	0.05	0.03
Salicylic acid	0.04	0.05	0.05	0.01	0.05	0.03	0.05	0.09
Tebuconazole	0.04	0.08	0.07	0.04	0.04	0.03	0.05	0.04
Tetraglyme	0.04	0.01	0.05	0.06	0.04	0.00	0.05	0.05
Thiacloprid	0.04	0.06	0.06	0.05	0.05	0.10	0.05	0.03
Triglyme	0.04	0.05	0.03	0.05	0.04	0.04	0.04	0.04

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-13 – Results of relative retention time (RRT) target analytes during optimisation of LC-column. In the table, the mean relative RT of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) is shown. The RRT shows the ratio of the target analyte's retention time to the retention time of ISTD (diuron-D6 in positive and mecoprop-D3 in negative ionisation mode). The results based on mobile phase 2 (see Table-A 3-8).

Target compounds	Kinetex		Synergi		Raptor Biphenyl		Ultra Aqueous	
	RRT	SD	RRT	SD	RRT	SD	RRT	SD
Iminopyrazole Acid-3	1.55	0.00	1.54	0.01	0.63	0.02	1.32	0.04
1,5-Diaminonaphthalene	0.14	0.07	0.17	0.08	0.25	0.04	0.21	0.00
2-(n-Ethylanilino)ethanol	0.26	0.07	0.3	0.03	0.38	0.06	0.31	0.03
TBSA	0.49	0.06	0.53	0.01	0.59	0.07	0.55	0.00
2-(Trifluoromethyl)benzamide	0.49	0.05	0.61	0.06	0.57	0.00	0.53	0.03
2,2'-(4-Methylphenylimino)diethanol	0.39	0.01	0.47	0.09	0.48	0.07	0.41	0.03
Melamine	0.39	0.08	0.47	0.05	0.09	0.03	0.09	0.04
AMTT	0.58	0.05	0.7	0.09	0.57	0.03	0.61	0.03
6-Chloropyridine-2-carboxylic acid	0.42	0.06	0.53	0.00	0.57	0.08	0.48	0.01
Acifluofen	1.1	0.08	1.1	0.03	1.13	0.01	1.08	0.04
Azoxystrobin	1.13	0.04	1.22	0.06	1.35	0.06	1.04	0.04
ABC 700	0.35	0.03	0.43	0.09	0.42	0.04	0.41	0.00
Benzamideoxime	0.14	0.07	0.15	0.06	0.18	0.08	0.19	0.00
Benzoic acid	0.16	0.01	0.24	0.01	0.18	0.05	0.19	0.00
Benzenesulfonic acid	0.36	0.03	0.41	0.09	0.43	0.03	0.41	0.03
Carbamazepine	0.94	0.07	1.04	0.06	1.08	0.07	0.92	0.02
Ciprofloxacin	0.51	0.01	0.52	0.02	0.63	0.00	0.55	0.01
Climbazole	0.94	0.02	0.99	0.08	0.99	0.01	0.86	0.05
Clothianidin	0.53	0.03	0.62	0.08	0.62	0.05	0.61	0.04
Cyproconazole	1.19	0.02	1.3	0.09	1.2	0.03	1.1	0.03
Diglyme	0.39	0.03	0.33	0.03	0.47	0.02	0.5	0.02
Diphenyl sulfone	0.92	0.03	1.03	0.08	1.11	0.07	0.87	0.04
Fluopyram	1.22	0.01	1.31	0.09	1.21	0.03	1.09	0.05
Imidacloprid	0.55	0.04	0.63	0.07	0.8	0.06	0.65	0.05
Lactofen	1.44	0.01	1.53	0.06	1.48	0.03	1.25	0.01
Metanilic acid	0.08	0.07	0.08	0.03	0.08	0.05	0.09	0.04
DEF 100	1.32	0.03	1.47	0.03	1.35	0.07	1.23	0.04
Morpholine	0.09	0.09	0.08	0.03	0.08	0.00	0.08	0.05
Moxifloxacin	0.67	0.06	0.69	0.08	0.84	0.07	0.69	0.01
N-(M-Tolyl)-diethanolamine	0.34	0.02	0.4	0.10	0.45	0.00	0.37	0.05
DEET	1.02	0.06	1.14	0.05	1.12	0.02	0.98	0.02
DMSA	0.73	0.01	0.84	0.02	0.85	0.02	0.7	0.03
N-Nitrosomorpholine	0.19	0.03	0.26	0.08	0.3	0.07	0.29	0.01
Phthalic acid	0.36	0.05	0.42	0.03	0.43	0.02	0.41	0.00
Picolinafen	1.45	0.07	1.54	0.01	1.44	0.03	1.29	0.03
Prothioconazole	1.3	0.03	1.4	0.09	1.31	0.02	1.16	0.01
Py-ethylamine	0.51	0.07	0.55	0.07	0.52	0.02	0.5	0.02
Salicylic acid	0.59	0.03	0.65	0.09	0.6	0.01	0.65	0.03
Tebuconazole	1.29	0.07	1.39	0.06	1.27	0.05	1.17	0.03
Tetraglyme	0.57	0.00	0.67	0.00	0.72	0.07	0.65	0.01
Thiacloprid	0.69	0.06	0.79	0.08	0.98	0.04	0.78	0.05
Triglyme	0.49	0.07	0.59	0.04	0.61	0.00	0.58	0.02



3.6.2.6 Comparison of Fragmentation Methods

Table-A 3-14 – Recovery of target analytes in industrial wastewater (SP 3) using DDA and DIA for LC-HR-MS/MS. The comparison was made using three replicates (Rep. 1 to Rep 3.). 'N.D.'  $\triangleq$  the target analyte was not detected.

Target compounds	DDA intensity in cps			vDIA intensity in cps		
	Rep. 1	Rep. 2	Rep. 3	Rep. 1	Rep. 2	Rep. 3
Iminopyrazole Acid-3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
1,5-Diaminonaphthalene	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
1,6-Naphthalenedisulfonic acid	1812	1612	1557	1566	1346	1616
1,7-Naphthalenedisulfonic acid	1812	1622	1692	1566	1268	1045
2-(n-Ethylanilino)ethanol	331100	345100	358100	287500	289100	272500
TBSA	217	N.D.	209	133	N.D.	N.D.
2-(Trifluoromethyl)benzamide	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2,2'-(4-Methylphenylimino)diethanol	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Melamine	2738	2648	2815	1775	1514	1591
2,4,6-Trimethoxy-1,3,5-triazine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
AMTT	N.D.	N.D.	101	N.D.	N.D.	N.D.
2-Chlorobenzoic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2-Naphthalenesulfonic acid	9894	9764	9911	8038	7915	7905
3,5-Dihydroxybenzoic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
3-Aminobenzenesulfonic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
3-Hydroxy-2-nitrobenzoic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
3-Nitrophthalic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
4-Aminobenzoic acid	1062	915	987	N.D.	N.D.	N.D.
CCIM	102700	110700	105500	98630	99930	96730
4-Chlorobenzoic acid	26240	28740	29740	20410	19910	19810
4-Hydroxyphenylglycolic acid	N.D.	115	N.D.	N.D.	N.D.	N.D.
4-Sulphophthalic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
4-Toluenesulfonic acid	203700	225700	215700	164900	160000	152900
6-Chloropyridine-2-carboxylic acid	2164	2264	2194	2217	2007	2127
Acifluorfen	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Azoxystrobin	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Azoxystrobin-D4	229600	231600	232600	204000	200000	198000
ABC 700	110200	112200	111200	95630	94630	95030
Benzamideoxime	269800	272800	271800	192300	190300	55200
Benzenesulfonic acid	79550	80150	80250	59860	59060	58860
Benzoic acid	290200	298200	292200	216500	212500	216500
Bezafibrate-D6	91530	83850	89850	80680	79090	79190
Carbamazepine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Chloramphenicol-D5	28570	29170	28970	24520	25520	19520
Chlorphenol (sum)	2296	2326	2306	1903	2003	1993
Ciprofloxacin	334	353	333	484	184	N.D.
Climbazole	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Clothianidin	827	697	817	380	300	279
Clothianidin-D3	18800	19800	18900	18530	18030	19130
Cyproconazole	1619	1729	1659	1336	1296	1316
Diclofenac-D4	34570	26140	29140	25970	24970	25270
Diglyme	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Diphenyl sulfone	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Ditolyetherdisulfonic acid	258	247	N.D.	200	N.D.	N.D.
Diuron	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Diuron-D6	61460	62560	61960	60580	59120	98580
Fluopyram	N.D.	N.D.	N.D.	163	N.D.	N.D.
Imidacloprid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Ircadine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Lactofen	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Maleic acid	68140	69240	68940	46910	50010	45810
Mecoprop-D3	25480	25480	26580	20280	21280	19980
Mesotrione-D4	1482	1592	1532	5675	5765	5635
Metsulfuron-methyl-D3	7812	7972	7672	6875	6745	6645
DEF 100	509	N.D.	109	543	N.D.	N.D.
Morpholine	21490	24190	23290	15460	15060	13460
Moxifloxacin	628	428	N.D.	549	549	N.D.
N-(M-Tolyl)-diethanolamine	7198	7538	7428	5733	5613	5423
DEET	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
DEET-D7	8447	8577	8607	7522	7212	7012
DMSA	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
N-Allylthiourea	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
N-Nitrosomorpholine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-14 continued.

N-Phenylurea	269800	277800	289800	192300	197800	180300
Phthalic acid	883900	899900	867900	658200	643200	600200
Phthalic anhydride	175100	183100	155100	216200	226200	196200
Phthalimide	14690	15590	16690	12470	16370	14270
Picolinafen	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Prothioconazole	1907	1807	1787	1506	1436	1126
Py-ethylamine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Salicylic acid	84630	84530	86730	70900	71200	69900
Salicylic acid-D4	13260	29260	28260	11580	11480	13080
Spiroxamine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Sulcotrione-D4	7480	7990	7510	5488	5328	5788
Tebuconazole	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Tebuconazole-D6	5658	5928	5678	5195	5005	5115
Tetraglyme	490	371	610	N.D.	586	786
Thiacloprid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Thiourea	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Triethylamine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Triglyme	64490	61390	67090	45970	43070	41270
Trimethylcarboxylic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

### 3.6.2.7 Peak Picking using Intensity Threshold

Table-A 3-15 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the intensity threshold of 100 cps. 'Identified' means the target was detected in all samples within the sampling site.

Target compounds	Mean of SP 1	Mean of SP 2	Mean of SP 3	Mean of SP 4	Mean of SP 5	Mean of UPW
Iminopyrazole Acid-3	identified	identified	identified	identified	identified	identified
1,5-Diaminonaphthalene	identified	identified	identified	identified	identified	identified
1,7-Naphthalenedisulfonic acid	identified	identified	identified	identified	identified	identified
2-(n-Ethylanilino)ethanol	identified	identified	identified	identified	identified	identified
TBSA	identified	identified	identified	identified	identified	identified
2-(Trifluoromethyl)benzamide	identified	identified	identified	identified	identified	identified
2,2'-(4-Methylphenylimino)diethanol	identified	identified	identified	identified	identified	identified
Melamine	identified	identified	identified	identified	identified	identified
2,4,6-Trimethoxy-1,3,5-triazine	identified	identified	identified	identified	identified	identified
AMTT	identified	identified	identified	identified	identified	identified
2-Chlorobenzoic acid	identified	identified	identified	identified	identified	identified
2-Naphthalenesulfonic acid	identified	identified	identified	identified	identified	identified
3-Hydroxy-2-nitrobenzoic acid	identified	identified	identified	identified	identified	identified
4-Aminobenzoic acid	identified	identified	identified	identified	identified	identified
CCIM	identified	identified	identified	identified	identified	identified
4-Chlorobenzoic acid	identified	identified	identified	identified	identified	identified
4-Toluolsulfonic acid	identified	identified	identified	identified	identified	identified
6-Chloropyridine-2-carboxylic acid	identified	identified	identified	identified	identified	identified
Aciflurofen	identified	identified	identified	identified	identified	identified
Azoxystrobin	identified	identified	identified	identified	identified	identified
ABC 700	identified	identified	identified	identified	identified	identified
Benzamideoxime	identified	identified	identified	identified	identified	identified
Benzenesulfonic acid	identified	identified	identified	identified	identified	identified
Carbamazepine	identified	identified	identified	identified	identified	identified
Chlorophenol (sum)	identified	identified	identified	identified	identified	identified
Ciprofloxacin	identified	identified	identified	identified	identified	identified
Climbazole	identified	identified	identified	identified	identified	identified
Clothianidin	identified	identified	identified	identified	identified	identified
Cyproconazole	identified	identified	identified	identified	identified	identified
Diglyme	identified	identified	identified	identified	identified	identified
Diphenyl sulfone	identified	identified	identified	identified	identified	identified
Ditolyletherdisulfonic acid	identified	identified	identified	identified	identified	identified
Fluopyram	identified	identified	identified	identified	identified	identified
Imidacloprid	identified	identified	identified	identified	identified	identified
Lactofen	identified	identified	identified	identified	identified	identified
Metanilic acid	identified	identified	identified	identified	identified	identified
DEF 100	identified	identified	identified	identified	identified	identified
Morpholine	identified	identified	identified	identified	identified	identified
Moxifloxacin	identified	identified	identified	identified	identified	identified

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-15 continued.

N-(M-Tolyl)-diethanolamine	identified	identified	identified	identified	identified	identified
DEET	identified	identified	identified	identified	identified	identified
DMSA	identified	identified	identified	identified	identified	identified
N-Nitrosomorpholine	identified	identified	identified	identified	identified	identified
Phthalic acid	identified	identified	identified	identified	identified	identified
Phthalimide	identified	identified	identified	identified	identified	identified
Picolinafen	identified	identified	identified	identified	identified	identified
Prothioconazole	identified	identified	identified	identified	identified	identified
Py-ethylamine	identified	identified	identified	identified	identified	identified
Salicylic acid	identified	identified	identified	identified	identified	identified
Saltidine	identified	identified	identified	identified	identified	identified
Spiroxamine	identified	identified	identified	identified	identified	identified
Tebuconazole	identified	identified	identified	identified	identified	identified
Tetraglyme	identified	identified	identified	identified	identified	identified
Thiacloprid	identified	identified	identified	identified	identified	identified
Triglyme	identified	identified	identified	identified	identified	identified

Table-A 3-16 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the intensity threshold of 500 cps. 'Identified' means the target was detected in all samples within the sampling site. 'N.D.' means the target was not detected in the industrial wastewater samples.

Target compounds	Mean of SP 1	Mean of SP 2	Mean of SP 3	Mean of SP 4	Mean of SP 5	Mean of UPW
Iminopyrazole Acid-3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
1,5-Diaminonaphthalene	identified	identified	identified	identified	identified	identified
1,7-Naphthalenedisulfonic acid	identified	identified	identified	identified	identified	identified
2-(n-Ethylanilino)ethanol	identified	identified	identified	identified	identified	identified
TBSA	identified	identified	identified	identified	identified	identified
2-(Trifluoromethyl)benzamide	identified	identified	identified	identified	identified	identified
2,2'-(4-Methylphenylimino)diethanol	identified	identified	identified	identified	identified	identified
Melamine	N.D.	N.D.	N.D.	identified	identified	identified
2,4,6-Trimethoxy-1,3,5-triazine	identified	identified	identified	identified	identified	identified
AMTT	identified	identified	identified	identified	identified	identified
2-Chlorobenzoic acid	identified	identified	identified	identified	identified	identified
2-Naphthalenesulfonic acid	identified	identified	identified	identified	identified	identified
3-Hydroxy-2-nitrobenzoic acid	identified	identified	identified	identified	identified	identified
4-Aminobenzoic acid	identified	identified	identified	identified	identified	identified
CCIM	N.D.	identified	identified	identified	identified	identified
4-Chlorobenzoic acid	identified	identified	identified	identified	identified	identified
4-Toluolsulfonic acid	identified	identified	identified	identified	identified	identified
6-Chloropyridine-2-carboxylic acid	identified	identified	identified	identified	identified	identified
Acifluofen	N.D.	N.D.	N.D.	identified	identified	identified
Azoxystrobin	identified	identified	identified	identified	identified	identified
ABC 700	identified	identified	identified	identified	identified	identified
Benzamideoxime	identified	identified	identified	identified	identified	identified
Benzenesulfonic acid	identified	identified	identified	identified	identified	identified
Carbamazepine	identified	identified	identified	identified	identified	identified
Chlorphenol (sum)	identified	identified	identified	identified	identified	identified
Ciprofloxacin	identified	identified	identified	identified	identified	identified
Climbazole	identified	identified	identified	identified	identified	identified
Clothianidin	identified	identified	identified	identified	identified	identified
Cyproconazole	identified	identified	identified	identified	identified	identified
Diglyme	identified	identified	identified	identified	identified	identified
Diphenyl sulfone	identified	identified	identified	identified	identified	identified
Ditolyetherdisulfonic acid	identified	identified	identified	identified	identified	identified
Fluopyram	identified	identified	identified	identified	identified	identified
Imidacloprid	identified	identified	identified	identified	identified	identified
Lactofen	identified	identified	identified	identified	identified	identified
Metanilic acid	N.D.	N.D.	N.D.	N.D.	N.D.	identified
DEF 100	identified	identified	identified	identified	identified	identified
Morpholine	N.D.	N.D.	N.D.	N.D.	identified	identified
Moxifloxacin	identified	identified	identified	identified	identified	identified
N-(M-Tolyl)-diethanolamine	identified	identified	identified	identified	identified	identified
DEET	identified	identified	identified	identified	identified	identified
DMSA	identified	identified	identified	identified	identified	identified
N-Nitrosomorpholine	identified	identified	identified	identified	identified	identified
Phthalic acid	identified	identified	identified	identified	identified	identified
Phthalimide	identified	identified	identified	identified	identified	identified
Picolinafen	identified	identified	identified	identified	identified	identified
Prothioconazole	identified	identified	identified	identified	identified	identified
Py-ethylamine	identified	identified	identified	identified	identified	identified

## Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-16 continued.

Salicylic acid	identified	identified	identified	identified	identified	identified
Saltidine	identified	N.D.	N.D.	identified	identified	identified
Spiroxamine	identified	identified	identified	identified	identified	identified
Tebuconazole	identified	identified	identified	identified	identified	identified
Tetraglyme	identified	identified	identified	identified	identified	identified
Thiacloprid	identified	identified	identified	identified	identified	identified
Triglyme	N.D.	N.D.	N.D.	N.D.	identified	identified

Table-A 3-17 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the threshold of 1,000 cps. 'Identified' means the target was detected in all samples within the sampling site. 'N.D.' means the target was not detected in the samples.

Target compounds	Mean of SP 1	Mean of SP 2	Mean of SP 3	Mean of SP 4	Mean of SP 5	Mean of UPW
Iminopyrazole Acid-3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
1,5-Diaminonaphthalene	N.D.	N.D.	N.D.	N.D.	identified	identified
1,7-Naphthalenedisulfonic acid	identified	identified	identified	identified	identified	identified
2-(n-Ethylanilino)ethanol	identified	identified	identified	identified	identified	identified
TBSA	identified	identified	identified	identified	identified	identified
2-(Trifluoromethyl)benzamide	identified	identified	identified	identified	identified	identified
2,2'-(4-Methylphenylimino)diethanol	identified	identified	identified	identified	identified	identified
Melamine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2.4.6-Trimethoxy-1,3,5-triazine	N.D.	identified	identified	identified	identified	identified
AMTT	identified	identified	identified	identified	identified	identified
2-Chlorobenzoic acid	N.D.	N.D.	identified	identified	identified	identified
2-Naphthalenesulfonic acid	identified	identified	identified	identified	identified	identified
3-Hydroxy-2-nitrobenzoic acid	identified	identified	identified	identified	identified	identified
4-Aminobenzoic acid	identified	identified	identified	identified	identified	identified
CCIM	identified	identified	identified	identified	identified	identified
4-Chlorobenzoic acid	identified	identified	identified	identified	identified	identified
4-Toluolsulfonic acid	identified	identified	identified	identified	identified	identified
6-Chloropyridine-2-carboxylic acid	identified	identified	identified	identified	identified	identified
Acifluorfen	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Azoxystrobin	identified	identified	identified	identified	identified	identified
ABC 700	identified	identified	identified	identified	identified	identified
Benzamidoxime	identified	identified	identified	identified	identified	identified
Benzenesulfonic acid	identified	identified	identified	identified	identified	identified
Carbamazepine	identified	identified	identified	identified	identified	identified
Chlorphenol (sum)	N.D.	N.D.	identified	identified	identified	identified
Ciprofloxacin	identified	identified	identified	identified	identified	identified
Climbazole	identified	identified	identified	identified	identified	identified
Clothianidin	identified	identified	identified	identified	identified	identified
Cyproconazole	identified	identified	identified	identified	identified	identified
Diglyme	identified	identified	identified	identified	identified	identified
Diphenyl sulfone	N.D.	N.D.	N.D.	identified	identified	identified
Ditolyetherdisulfonic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Fluopyram	identified	identified	identified	identified	identified	identified
Imidacloprid	identified	identified	identified	identified	identified	identified
Lactofen	N.D.	N.D.	N.D.	identified	identified	identified
Metanilic acid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
DEF 100	identified	identified	identified	identified	identified	identified
Morpholine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Moxifloxacin	identified	identified	identified	identified	identified	identified
N-(M-Tolyl)-diethanolamine	identified	identified	identified	identified	identified	identified
DEET	identified	identified	identified	identified	identified	identified
DMSA	identified	identified	identified	identified	identified	identified
N-Nitrosomorpholine	identified	identified	identified	identified	identified	identified
Phthalic acid	identified	identified	identified	identified	identified	identified
Phthalimide	identified	identified	identified	identified	identified	identified
Picolinafen	identified	identified	identified	identified	identified	identified
Prothioconazole	identified	identified	identified	identified	identified	identified
Py-ethylamine	identified	identified	identified	identified	identified	identified
Salicylic acid	identified	identified	identified	identified	identified	identified
Saltidine	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Spiroxamine	N.D.	N.D.	identified	identified	identified	identified
Tebuconazole	identified	identified	identified	identified	identified	identified
Tetraglyme	identified	identified	identified	identified	identified	identified
Thiacloprid	identified	identified	identified	identified	identified	identified
Triglyme	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

3.6.2.8 Retention Time and Mass Error

Table-A 3-18 – Results of Retention time and mass error tolerance. RT  $\pm$  Retention time.

ID	Target compounds	RT in min	Tolerance of RT in %	Tolerance of RT in min	m/z [ESI <sup>+</sup> ]	m/z [ESI <sup>-</sup> ]	Mass error tolerance in ppm
1	Iminopyrazole Acid-3	2.46	-0.24	-0.01	-	252.045	2.3
2	1,5-Diaminonaphthalene	1.94	0.08	0.00	159.092	-	2.4
3	1,6-Naphthalenedisulfonic acid	2.03	-0.19	0.00	-	286.969	2.5
4	1,7-Naphthalenedisulfonic acid	2.03	-0.19	0.00	-	286.969	2.5
5	2-(n-Ethylanilino)ethanol	2.68	0.49	0.01	166.123	-	-1
6	TBSA	4.96	-0.14	-0.01	-	224.000	0.2
7	2-(Trifluoromethyl)benzamide	5.32	0.69	0.04	190.047	-	-0.9
8	2,2'-(4-Methylphenylimino)diethanol	3.25	0.25	0.01	196.133	-	2.4
9	Melamine	0.73	0.47	0.00	127.073	-	1.9
10	2,4,6-Trimethoxy-1,3,5-triazine	6.77	0.65	0.04	172.072	-	2.1
11	AMTT	5.9	0.82	0.05	195.049	-	-0.6
12	2-Chlorobenzoic acid	6.72	0.05	0.00	157.005	-	-0.6
13	2-Naphthalenesulfonic acid	5.29	-0.17	-0.01	-	207.012	0
14	3,5-Dihydroxybenzoic acid	2.66	-0.02	0.00	-	153.093	-0.7
15	3-Aminobenzenesulfonic acid	0.71	-0.14	0.00	-	172.007	0
16	3-Hydroxy-2-nitrobenzoic acid	3.44	-0.15	-0.01	-	182.009	0.4
17	3-Nitrophthalic acid	2.66	-0.32	-0.01	-	210.004	-0.5
18	4-Aminobenzoic acid	2.8	0.41	0.01	138.055	-	-0.6
19	CCIM	13.06	0.46	0.06	218.048	-	-0.1
20	4-Chlorobenzoic acid	9.77	-0.13	-0.01	-	154.991	-0.4
21	4-Hydroxyphenylglycolic acid	1.56	-0.01	0.00	-	167.035	-0.9
22	4-Sulphophthalic acid	0.97	0.59	0.01	-	244.976	-0.7
23	4-Toluenesulfonic acid	2.82	-0.08	0.00	-	171.012	0.3
24	6-Chloropyridine-2-carboxylic acid	4.3	0.92	0.04	158.000	-	-0.7
25	Acifluorfen	13.77	-0.25	-0.03	-	359.989	-0.6
26	Azoxystrobin	13.07	0.42	0.05	404.124	-	-0.6
27	Azoxystrobin-D4	13.04	0.28	0.04	408.149	-	-0.2
28	ABC 700	3.56	0.06	0.00	177.047	-	-0.9
29	Benzamideoxime	1.78	1.10	0.02	137.071	-	2.3
30	Benzenesulfonic acid	1.79	-0.15	0.00	-	156.996	3.1
31	Benzoic acid	3.66	0.12	0.00	-	121.030	3.5
32	Bezafibrate-D6	12.62	0.39/-0.44	0.05/-0.06	368.153	366.138	-0.1
33	Carbamazepine	10.84	0.44	0.05	237.102	-	-0.8
34	Chloramphenicol-D5	6.12	-0.20	-0.01	-	326.036	-0.2
35	Chlorphenol (sum)	7.75	-0.08	-0.01	-	126.996	2.1
36	Ciprofloxacin	5.7	0.83	0.05	332.140	-	3
37	Climbazole	10.14	0.06	0.01	293.105	-	2.9
38	Clothianidin	6.15	0.43	0.03	250.016	-	1
39	Clothianidin-D3	6.12	0.52	0.03	253.035	-	0.4
40	Cyproconazol	13.73	0.34	0.05	292.121	-	0.5
41	Diclofenac-D4	14.11	0.39/-0.23	0.06/-0.04	300.049	298.035	4.1
42	Diglyme	4.62	0.70	0.03	135.102	-	0.8
43	Diphenyl sulfone	10.15	0.02	0.00	236.074	-	0.2
44	Ditolyletherdisulfonic acid	3.97	0.15	0.01	-	357.011	0.9
45	Diuron	11.98	0.47	0.06	233.024	-	1.7
46	Diuron-D6	11.95 / 11.93	0.49/-0.06	0.07/0.01	239.062	237.047	1.6
47	Fluopyram	13.82	0.32	0.04	397.054	-	1.3
48	Imidacloprid	6.7	0.59	0.04	256.060	-	4.8
49	Ircadine	12.12	0.47	0.06	230.175	-	3.6
50	Lactofen	16.53	0.34	0.06	479.080	-	4.1
51	Maleic acid	12.23	-0.15	-0.02	-	115.004	1
52	Mecoprop-D3	8.52	-0.06	0.00	-	216.051	0.8
53	Mesotrione-D4	10.18 / 10.17	0.37/0.32	0.04/0.03	344.074	342.059	0.5
54	Metsulfuron-methyl-D3	15.65	0.47/0.07	0.06/0.01	385.100	383.086	4.7
55	DEF 100	0.64	0.37	0.00	229.043	-	4.6
56	Morpholine	7.7	0.35	0.03	88.076	-	4.6
57	Moxifloxacin	3.72	0.56	0.02	402.182	-	5
58	N-(M-Tolyl)-diethanolamine	11.67	0.39	0.05	196.133	-	3.1
59	DEET	11.6	0.57	0.07	192.138	-	4.8
60	DEET-D7	7.42	0.47	0.03	199.182	-	4.9
61	DMSA	2.08	0.56	0.01	201.069	-	4.6
62	N-Allylthiourea	2.51	0.78	0.02	117.048	-	3.1
63	N-Nitrosomorpholine	3.84	0.03	0.00	117.066	-	-0.8
64	N-Phenylurea	3.67	1.10	0.04	137.071	-	-0.9
65	Phthalic acid	15.98	-0.04	-0.01	-	165.019	0.8

Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-18 continued.

66	Phthalic anhydride	5.43	0.31	0.02	149.023	-	-1
67	Phthalimide	17.02	0.71	0.02	148.039	-	-0.1
68	Picolinafen	14.74	0.24	0.04	377.091	-	-0.1
69	Prothioconazole	4.51	0.20	0.01	344.039	-	0
70	Py-ethylamine	6.29	1.34	0.08	225.040	-	-0.5
71	Salicylic acid	6.22	-0.14	-0.01	-	137.024	-0.5
72	Salicylic acid-D4	12	-0.08	-0.01	-	141.050	-0.8
73	Spiroxamine	9.71	0.03	0.00	298.274	-	-0.3
74	Sulcotrione-D4	14.94	0.02/-0.25	0.00/-0.03	333.050	331.035	0
75	Tebuconazole	14.91	0.35	0.05	308.152	-	0
76	Tebuconazole-D6	6.82	0.36	0.02	314.190	-	0.4
77	Tetraglyme	8.63	0.68	0.06	223.154	-	0.3
78	Thiacloprid	0.71	0.51	0.00	253.031	-	-0.5
79	Thiourea	1.88	0.21	0.00	77.017	-	4.2
80	Triethylamine	5.8	0.35	0.02	102.128	-	3.3
81	Triglyme	8.8	0.47	0.04	179.128	-	2.9
82	Trimethylcarboxylic acid	8.8	0.10	0.01	204.102	-	-0.7

### 3.6.2.9 Validation of Non-Target Screening Method

Table-A 3-19 – Result of method validation of suspected-targets of 100 un-spiked wastewater sample (SP 5) in positive (+) and negative (-) ionisation mode. Previously, the samples were analysed with LC-UV declaring them positive for at least one suspected-target analyte.

<i>Ionisation</i>	<i>Component Name</i>	<i>The relative standard deviation of retention time in %</i>	<i>Max mass error tolerance in ppm</i>	<i>Min mass error tolerance in ppm</i>	<i>Max mass error tolerance in mDa</i>	<i>Min mass error tolerance in mDa</i>
(+)	1,5-Diaminonaphthalene	0.84	53.2	-4.9	8.47	-0.77
	2-(n-Ethylanilino)ethanol	0.86	0.8	-1	0.14	-0.17
	2-(Trifluoromethyl)benzamide	1.59	1.9	-0.9	0.36	-0.18
	2,2'-(4-Methylphenylimino)diethanol	1.04	0.8	-1	0.16	-0.19
	Melamine	0.07	20.5	-1.1	2.61	-0.15
	AMTT	1.89	0.9	-1	0.18	-0.19
	2-Chlorobenzoic acid	2.16	3.6	-1.9	0.56	-0.3
	4-Aminobenzoic acid	0.89	2.8	-1.7	0.39	-0.23
	CCIM	4.19	1.1	-0.7	0.25	-0.15
	6-Chloropyridine-2-carboxylic acid	1.38	4.5	-2.2	0.72	-0.35
	Azoxystrobin	4.19	1.2	-1	0.49	-0.4
	ABC 700	1.14	3.5	-0.8	0.61	-0.14
	Benzamideoxime	0.84	1	-86	0.14	-11.79
	Carbamazepine	1.94	44.9	-1	10.65	-0.24
	Ciprofloxacin	1.84	2.3	-0.4	0.77	-0.14
	Climbazole	3.23	1	-1	0.29	-0.29
	Clothianidin	1.97	4.4	-0.4	1.09	-0.11
	Cyproconazole	4.4	1	-0.8	0.29	-0.23
	Diglyme	1.27	115.3	-3.5	15.58	-0.47
	Diphenyl sulfone	3.25	4.7	-0.9	1.12	-0.22
	Diuron	3.84	1	-0.9	0.23	-0.21
	Fluopyram	4.43	1.5	-1	0.61	-0.39
	Imidacloprid	2.15	3.5	-0.5	0.89	-0.12
	Lactofen	5.3	2	0.1	0.94	0.05
	DEF 100	5.02	3	-1	0.68	-0.23
	Morpholine	0.17	2.2	-7.4	0.19	-0.65
	Moxifloxacin	2.5	2.6	0.1	1.06	0.06
	N-(M-Tolyl)-diethanolamine	1.18	1	-1	0.19	-0.2
	DEET	3.74	0.5	-1	0.11	-0.19
	DMSA	1.85	100.3	-0.6	20.16	-0.11
	N-Allylthiourea	0.72	155.6	-4.6	18.21	-0.54
	N-Nitrosomorpholine	0.81	3.9	-2.1	0.45	-0.25
	N-Phenylurea	0.84	1	-86	0.14	-11.79
Phthalic anhydride	0.01	-0.3	-2.1	-0.04	-0.31	
Phthalimide	1.74	5.1	-5.4	0.76	-0.8	
Picolinafen	3.05	6.7	-4.7	2.53	-1.76	
Prothioconazole	0.05	1.3	-12.1	0.46	-4.17	

Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-19 continued.

	Py-ethylamine	1.13	16.2	-0.9	3.65	-0.21
	Saltidine	3.63	10.7	-0.9	2.45	-0.22
	Spiroxamine	3.86	1	-1	0.29	-0.3
	Tebuconazole	4.79	1.1	-1	0.33	-0.31
	Tetraglyme	1.71	69.7	-0.9	15.55	-0.21
	Thiacloprid	0.09	0.5	-132.5	0.12	-33.52
	Thiourea	0.1	3.7	-49.4	0.28	-3.8
	Triethylamine	0.24	1.8	-2.4	0.19	-0.24
	Triglyme	0.74	88.8	-1	15.91	-0.18
	Trimethylcarboxylic acid	2.8	2.2	-1	0.44	-0.2
	Iminopyrazole Acid-3	0.78	3.5	-2.6	0.88	-0.66
	1,6-Naphthalenedisulfonic acid	0.65	3.2	-2.1	0.93	-0.6
	1,7-Naphthalenedisulfonic acid	0.65	3.2	-2.1	0.93	-0.6
	TBSA	1.71	2.3	-2	0.52	-0.45
	2-Naphthalenesulfonic acid	1.57	1	-3.4	0.2	-0.71
	3,5-Dihydroxybenzoic acid	0.85	4.2	-3.9	0.64	-0.6
	3-Hydroxy-2-nitrobenzoic acid	1.11	3.3	-3.7	0.59	-0.67
	3-Nitrophthalic acid	1.34	8.2	-7.3	1.73	-1.52
	4-Chlorobenzoic acid	3.13	5	-2.7	0.77	-0.43
	4-Hydroxyphenylglycolic acid	0.74	3.7	-9.3	0.62	-1.56
(-)	4-Sulphophthalic acid	0.47	5.6	-6.5	1.38	-1.6
	4-Toluenesulfonic acid	0.9	0.9	-1	0.15	-0.17
	Acifluorfen	4.42	1.6	-1.9	0.58	-0.7
	Benzenesulfonic acid	0.57	2.8	-1	0.45	-0.15
	Benzoic acid	1.37	7.4	-6.2	0.9	-0.75
	Chlorphenol (sum)	2.76	13.8	-4.2	1.75	-0.54
	Ditolyetherdisulfonic acid	1.27	6.5	-3.7	2.34	-1.32
	Maleic acid	0.54	14.8	-8.2	1.7	-0.94
	Metanilic acid	0.23	2.7	-1.5	0.47	-0.25
	Phthalic acid	1.18	3.8	-2.3	0.62	-0.37
	Salicylic acid	2.03	5.9	-1.3	0.81	-0.18

3.6.2.10 Application to authentic standards

Table-A 3-20 – Exemplary feature table of treated wastewater (SP 5) in positive ionisation mode (+).

<i>m/z</i>	<i>log10 sum intensity</i>	<i>RT in min</i>	<i>In replicates?</i>
475.4137	4.81312	23.81	TRUE
548.50247	4.89771	23.82	TRUE
663.4525	4.43805	22.02	TRUE
419.35157	4.44848	23.81	TRUE
549.50666	4.49416	23.78	TRUE
476.41778	4.34908	23.80	TRUE
553.45874	4.51359	23.79	TRUE
664.45672	4.08262	22.02	TRUE
492.4403	4.08403	22.15	TRUE
647.45753	4.23379	24.80	TRUE
475.41389	4.04418	22.15	TRUE
312.36205	3.94765	18.44	TRUE
408.14836	3.67846	13.00	TRUE
406.32893	3.82591	19.9	TRUE
420.35532	3.91423	23.81	TRUE
675.57681	4.22127	19.74	FALSE
554.46262	4.10416	23.81	TRUE
384.34701	3.68261	19.91	TRUE
722.52687	3.94366	22.02	TRUE
531.47695	3.83927	23.78	TRUE
648.46174	3.90432	24.79	TRUE
338.3416	3.58353	19.91	TRUE
332.33052	3.54064	16.34	TRUE
497.39633	3.73325	22.16	TRUE
419.35163	3.67733	22.16	TRUE
685.43586	3.58009	22.02	TRUE
493.44437	3.62014	22.15	TRUE
550.51031	3.77976	23.78	TRUE
364.32389	4.0178	19.86	FALSE
676.58003	3.94873	19.74	TRUE
369.09231	3.34858	6.23	TRUE
476.4179	3.53282	22.15	TRUE
394.34651	3.35716	20.95	TRUE

Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-20 continued.

333.04915	3.6554	9.69	TRUE
723.53031	3.60851	22.03	TRUE
665.46045	3.46291	22.02	TRUE
477.42113	3.5707	23.79	TRUE
368.15233	3.28167	12.57	TRUE
383.29502	3.87689	20.42	TRUE
355.26393	3.82233	19.59	TRUE
338.34165	3.37804	20.82	TRUE
251.0658	3.29514	10.16	TRUE
532.48053	3.42691	23.78	TRUE
680.48039	3.3119	22.03	TRUE
341.09728	3.37321	7.39	TRUE
341.09727	3.15875	8.16	TRUE
391.28433	3.19219	19.86	TRUE
313.36583	3.29193	18.44	TRUE
284.33096	3.24019	17.35	TRUE
686.4394	3.2407	22.04	TRUE
885.36812	3.01097	19.88	TRUE
409.15235	3.08749	13.00	TRUE
407.33263	3.17956	19.9	TRUE
279.15899	3.09688	15.91	TRUE
149.02314	3.12877	15.92	TRUE
498.4001	3.26135	22.16	TRUE
385.35063	3.01492	19.91	TRUE
310.2377	3.25486	16.49	TRUE
301.14091	3.08853	15.92	TRUE
708.51167	3.21548	22.00	TRUE
304.29972	3.03795	15.00	TRUE
420.35535	3.15414	22.16	TRUE
149.04459	2.87837	1.67	TRUE
403.23245	2.99069	16.75	TRUE
335.04678	3.26836	9.68	TRUE
333.33444	2.96467	16.34	TRUE
657.5668	3.38022	19.78	FALSE
251.06611	2.98983	9.69	TRUE
385.10039	2.94536	10.14	TRUE
177.04676	2.89227	3.57	TRUE
419.35188	3.04148	21.22	TRUE
109.07568	2.87127	2.86	TRUE
489.31325	2.89596	20.06	TRUE
649.46589	3.28075	24.78	TRUE
141.06549	2.89652	3.54	TRUE
339.34521	2.97869	19.9	TRUE
555.46628	3.31794	23.83	TRUE
421.35852	3.1018	23.81	TRUE
447.34699	3.09038	20.96	TRUE
391.07482	2.8917	6.23	TRUE
279.09319	3.2118	13.15	TRUE
350.11311	2.8333	7.20	TRUE
370.1504	2.84165	12.57	TRUE
299.29447	2.7795	20.42	TRUE
430.13118	2.85446	13.00	TRUE
300.04918	2.84575	14.06	TRUE
681.48357	2.95014	22.01	TRUE
529.46168	3.2058	22.40	TRUE
663.56418	2.56998	19.55	FALSE
159.09141	2.82397	5.03	TRUE
236.05518	2.8459	5.33	TRUE
395.35039	2.81561	20.95	TRUE
237.04235	2.94379	11.36	TRUE
886.37059	2.76003	19.88	TRUE
724.53406	3.00482	22.03	TRUE
237.08657	2.83203	5.99	TRUE
239.06174	2.89433	11.89	TRUE
789.66878	2.66032	19.91	TRUE
352.23294	2.69189	4.40	TRUE
371.3157	2.89982	19.84	TRUE
308.20663	2.69731	3.73	TRUE
386.10588	2.76567	10.14	TRUE
109.07573	2.6792	2.67	TRUE
379.34256	3.40521	20.34	FALSE
228.19589	2.93601	13.03	TRUE
569.43343	3.23178	23.80	TRUE
363.0798	2.7952	8.16	TRUE
709.515	2.91343	22.01	TRUE



Chapter 3 Development of a Multicomponent LC-ESI-qTOF-MS Screening Method and Data Processing Strategy for Non-Target Screening in Industrial Wastewater

Table-A 3-20 continued.

167.07038	2.98996	23.81	TRUE
407.14312	2.73862	13.11	TRUE
350.07628	3.03735	9.66	TRUE
127.07247	2.54694	0.79	TRUE
409.13944	2.72098	14.02	TRUE
363.07976	2.93864	7.39	TRUE
334.05316	3.02471	9.68	TRUE
413.26625	2.71096	19.85	TRUE
205.08598	2.75598	15.92	TRUE
376.12308	2.69642	13.00	TRUE
409.1395	2.65021	14.14	TRUE
371.31926	2.9792	18.24	TRUE
84.05546	2.49708	2.29	TRUE
295.0922	2.71144	6.57	TRUE
476.41817	2.91889	23.38	TRUE
704.61101	2.52928	19.77	FALSE
294.15477	2.43862	2.30	TRUE
369.09296	2.64069	7.02	TRUE
302.04628	2.66963	14.06	TRUE
369.15616	2.65012	12.57	TRUE
250.1284	2.39349	1.91	TRUE
494.44772	2.82202	22.17	TRUE
194.13895	2.19574	0.66	TRUE
370.09613	2.62473	6.23	TRUE
113.10705	2.60771	1.52	TRUE
548.39034	2.61236	20.07	TRUE
551.51309	2.60037	23.78	TRUE
477.42141	2.7605	22.14	TRUE
241.0591	2.70599	11.89	TRUE
907.34983	2.58135	19.89	TRUE
145.10105	2.59218	5.83	TRUE
633.40319	2.60971	23.80	TRUE
392.28779	2.51857	19.85	TRUE
390.13503	2.59806	12.57	TRUE
425.21471	2.62102	16.76	TRUE
343.29677	2.78967	16.22	TRUE
790.67215	2.39404	19.92	TRUE
353.14541	2.86066	10.48	TRUE
551.51375	2.55622	23.80	TRUE
459.30229	2.43415	20.06	TRUE
527.33724	2.574	18.95	TRUE
666.46411	2.64502	22.03	TRUE
339.34516	2.74637	20.83	TRUE
436.37864	2.5709	21.21	TRUE
295.09253	2.54978	7.25	TRUE
533.48422	2.36383	23.78	TRUE
164.9206	2.78707	0.66	TRUE
393.29731	2.61759	19.83	TRUE
296.22074	2.75206	15.05	TRUE
431.25019	2.84373	20.04	TRUE
687.44311	2.60641	22.04	TRUE
344.07374	2.84071	8.53	TRUE
107.04917	2.45912	23.79	FALSE
469.32904	2.47695	20.96	TRUE
407.08249	2.54822	10.16	TRUE
254.06604	2.53953	5.33	TRUE
631.40199	2.53402	23.79	TRUE
234.08728	2.54566	5.73	TRUE
531.47694	2.56389	24.07	TRUE
305.24732	2.69015	19.39	TRUE
569.45242	2.72485	22.39	TRUE
478.42473	2.56583	23.81	TRUE
264.18062	2.37584	2.99	TRUE
633.40369	2.52613	23.84	TRUE
701.40976	2.55899	22.04	TRUE
190.04746	2.52029	4.94	TRUE
236.14916	2.14383	1.90	TRUE
149.05976	2.09048	23.84	TRUE
75.02589	2.43061	1.67	TRUE
490.31652	2.36841	20.07	TRUE
273.0485	2.58183	10.20	TRUE
475.36024	2.45555	23.80	TRUE
419.31578	2.9718	20.52	TRUE
908.35307	2.36496	19.89	TRUE
607.39131	2.52251	22.02	TRUE

Table-A 3-20 continued.

149.05973	2.42087	23.79	TRUE
520.43615	2.54632	22.29	TRUE
605.09092	2.46859	9.68	TRUE
569.4335	2.66005	24.01	TRUE
320.12179	2.47717	7.61	TRUE
610.18597	2.28855	23.81	FALSE
282.2792	2.4593	19.29	TRUE
530.46544	2.79815	22.43	TRUE
298.34682	2.53537	17.47	TRUE
615.42998	2.5572	23.76	TRUE
355.11331	2.41038	6.57	TRUE
170.04121	2.45739	4.94	TRUE
317.07466	2.43938	10.93	TRUE
285.33474	2.51607	17.35	TRUE
276.04801	2.42114	5.33	TRUE
547.4725	2.53903	22.32	TRUE
342.10109	2.54907	7.39	TRUE

## Chapter 4 Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches

*Submitted to Anal. Chem.: Purschke, K., Vosough, M., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Evaluation of non-target long-term LC-HRMS time series data using multivariate statistical approaches.*

### 4.1 Introduction

Proper operation of wastewater treatment plants is required to prevent the entry of pharmaceutical residues, biocides, pesticides and other chemicals from industry, so-called trace organic compounds (TrOC), into surface waters. Besides, knowledge of wastewater quality, monitoring of production changes and registered substances is a prerequisite for ensuring compliance with regulatory environmental safety standards. Routine monitoring programs for the analysis of industrial wastewater are based on target analysis approaches. They offer selectivity and sensitivity towards TrOC but are limited to measuring preselected and known compounds of interest. Therefore, other potentially present contaminants, not included in the target list, are not detected, even if they occur in high concentrations in the sample. To counteract this fact, the application of high-resolution mass spectrometers coupled to liquid chromatography (LC-HRMS) has increased in recent years (Brüggen and Schmitz, 2018; Hedgespeth et al., 2019; Hollender et al., 2017; Kiefer et al., 2019; Schmidt, 2018). Non-target screening (NTS), based on LC-HRMS, enables the detection of previously undetected organic compounds. The monitoring of wastewater streams comprehensively assesses water treatment processes and identifies contaminants at environmentally relevant concentrations (Hollender et al., 2017; Peter et al., 2019). Considering this, NTS methods based on LC-HRMS analysis became an analytical technique of increasing importance for the water monitoring of TrOC monitoring in recent years (Bader et al., 2016; Hollender et al., 2017; Schmidt, 2018).

For NTS, all substances retained chromatographically by the used LC method and ionised by electrospray ionisation are recorded (Peter et al., 2019). As demonstrated in other studies (Nürnberg et al., 2015), a proper pre-processing workflow must be applied to obtain reliable data. Generally, the following steps are included: (1) peak picking, where distinct chromatographic-mass peaks are extracted as features (mass-to-charge ratio, retention time, intensity), (2) blank subtraction, where features belonging to the matrix, contamination or the background are removed from the sample data, (3) feature alignment, where detected features

are linked across several samples and (4) componentisation, where features belonging to the same compounds (e.g. isotopes and adducts) are grouped. Componentisation is an essential but complex step during NTS. Features belonging to one compound provide complementary information about the molecular formula, which is, in turn, necessary for structure elucidation (Ruttkies et al., 2019; Schollée et al., 2016). Grouping of isotopes, adducts and in-source fragments must, therefore, carefully be considered for later data analysis.

However, due to the complexity of environmental samples and despite these data reduction efforts, thousands of LC-HRMS features are generated in NTS. Thus, the peak list generated by the NTS methods needs to be reduced to a list containing features of potential interest. Feature prioritisation can help to reduce the complexity of the data, enabling the interpretation and the identification of the most relevant TrOCs (Alygizakis et al., 2019). Different data-driven and experiment-driven prioritisation approaches are discussed in the literature (Hollender et al., 2017). In the area of wastewater treatment, prioritisation approaches based on principal component analysis (PCA) combined with the F-ratio method and partial least square linear-discriminant analysis (PLS-DA) coupled with univariate statistics have recently been reported (Hohrenk et al., 2019; Samanipour et al., 2017).

Trend analysis using multivariate chemometric algorithms of time series presents a complementary approach of prioritising relevant features with temporal trends of interest. Few chemometric-based studies have been developed for NTS of time series LC-HRMS data and feature prioritisation approaches. In this regard, data reduction strategies were performed by Spearman's rank correlation coefficient, hierarchical cluster analysis (HCA), Multivariate empirical bayes approach (MEBA) and Hotelling's  $T^2$  coefficient (Albergamo et al., 2019; Alygizakis et al., 2019; Chiaia-Hernández et al., 2017; Plassmann et al., 2016; Schollée et al., 2016). However, in contrast to the mentioned studies, features with increasing/decreasing intensity trends over time were of special interest in this study. Highly fluctuating features are no relevant TrOCs for industrial wastewater because these fluctuations describe the normal production processes. In contrast, the increasing/decreasing intensity trend over time indicate changes in the sewage system, which are exceptional and of importance. Features whose intensity change continuously over time, for example, show changes in production, which remained unknown by performing target analysis. By focussing on such features, hidden correlations and trends get manageable, which were so far unknown. In turn, this enables meaningful monitoring and treatment in wastewater treatment plants (WWTPs).

The data reduction strategy performed by Plassmann et al. (2016) introduced already prioritisation of features with increasing intensities. However, in that study, univariate statistics

were used. For complex industrial wastewater long-term time studies, the application of just one variable to describe the process is not possible. While univariate methods such as the Spearman's rank correlation (Plassmann et al., 2016) or Mann-Kendall test (Lamchin et al., 2019; Mondal et al., 2012; Sharma et al., 2019) are instrumental in identifying individual feature changes, they can yield insufficient results and performance for describing complex relations in data sets. Thus, to build predictive models based on multiple features describing the process that can improve performance, multivariate statistical methods need to be employed (Gowda et al., 2017; Huynh et al., 2017; Jalil and Rao, 2019).

Therefore, the main objective of this study was the implementation of multivariate chemometric algorithms for the comprehensive evaluation of long-term time trends prioritising features for identification in industrial wastewater having a significant increase/decrease in their intensity over time. Two complementary strategies were used: PCA and group-wise principal component analysis (GPCA), which is a sparse factorisation based on PCA (José Camacho et al., 2017), as well as a univariate statistical test (Spearman rank's correlation). Furthermore, *Tscore* ranking based on Hotelling's  $T^2$  statistic (D-statistic) and Q-statistic was evaluated for prioritising features and sampling times describing most variance of the PCA model (José Camacho et al., 2019). To the best of our knowledge, the combination of PCA, GPCA and univariate analysis describes the first application for multivariate statistical prioritisation of long-term time trends detecting continuously increasing/decreasing peak intensities, reducing the number of features selected for identification of TrOCs in industrial wastewater.

## 4.2 Experimental Section

### 4.2.1 Solvents and Chemicals

All solvents used in the present work were of LCMS grade. Methanol was from Honeywell Riedel-de-Haën<sup>TM</sup> (Seelze, Germany). Ultra-pure water was used from a Milli-Q<sup>®</sup> (Q-PoD<sup>®</sup>) water system from Merck KGaA (Darmstadt, Germany) and formic acid (99%) was purchased from Fisher Chemical (Geel, Belgium). All internal standards (ISTD) and target analytes were of a quality grade suitable for trace analysis, purity  $\geq 95\%$ . The corresponding chemical abstract service (CAS) numbers of used targets and ISTD are listed in the supporting information (see Table-A 4-2 and Table-A 4-3).

### 4.2.2 Sample Material

During the study, two different data sets were used. A complete description is listed in the SI (see Table-A 4-1). For the evaluation and validation of the multivariate time trend analysis, 29

Milli-Q® water samples spiked with 19 target substances were used. The pure water was spiked with targets at different concentrations simulating relevant production changes and unimportant time trends (see Table-A 4-2), with concentrations between 0.2 and 200 µg L<sup>-1</sup>. Relevant trends describe continuously rising and falling intensities of features over time, whereas unimportant time trends were simulated by fluctuation, straight-line and peak trends.

Additionally, a data set of 69 authentic industrial wastewater samples, from November 2018 to March 2019, of 24-h composite flow WWTP influent samples were used. The WWTP influent samples were collected in pre-cleaned 250-mL high-density polyethylene (HDPE) bottles until analysis, which took place immediately after the end of the sampling.

### 4.2.3 LC-HRMS Screening Analysis

The LC-HRMS analysis method employed was based on Purschke et al. (2020) (Purschke et al., 2020). The detailed description of the LC-HRMS acquisition method can be found in the Appendix (see Table-A 4-4). In brief, samples were diluted by a factor of 10 with Milli-Q® and 0.1% formic acid and were spiked with 10 µL ISTD (cf. Table-A 4-2). 5 µL of the prepared sample was injected. The chromatographic separation was performed using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) with Raptor™ Ultra Aqueous C18 analytical column (100 mm x 2.1 mm, 3.0 µm particle size) linked to Restek™ Trident cartridge (10 x 2.1 mm) and a filter (2 mm, 0.5 µm; Restek GmbH, Bad Homburg v. d. Höhe, Germany). For detection, samples were measured (in sequence) in positive followed by negative electrospray ionisation mode (Turbo V™ of SCIEX GmbH, Darmstadt, Germany) on a hybrid quadrupole time-of-flight mass spectrometer x500R qTOF (SCIEX GmbH, Darmstadt, Germany). Full scan HRMS data were recorded within an *m/z* range of 70 – 950 for each sample with ten data-dependent fragmentation experiments (DDA, *m/z* 30 – 700).

### 4.2.4 Quality Control

To avoid false positive or false negative detection of trends resulting from instrumental drift, the stability/accuracy of the used LC-HRMS system should be monitored by analysing quality control (QC) samples repeatedly throughout the sample sequence. The QC standard consisted of 70 target analytes (see Table-A 4-5) with 100 µg L<sup>-1</sup> concentration and an internal standard mixture of 13 isotope-labelled (deuterated) compounds (for concentrations see Table-A 4-3). The QC standard was run within each batch for both ionisation modes at the beginning and the end of each sequence. Furthermore, within each batch, a blank, consisting of 1 mL Milli-Q® and 0.1% formic acid spiked with ISTD, was measured to monitor for background contaminations. Furthermore, the (pre-)column of the LC system was changed depending on

the peak shape of the analytes of the quality control. For internal and external standards, if the mass error exceeded 5 ppm, the instrument was re-tuned and samples were re-analysed. Additionally, after sample measurement, the ISTDs were analysed first. ISTDs should be categorised in the class of non-relevant trends (no trend), as they were spiked with the same concentration in each sample. An overview of detection, mass deviation and intensity ranges of the ISTDs is given in 4.6.3.1.1 (see Table-A 4-6). As expected, the ISTDs showed no relevant time trend but showed a constant intensity over time.

#### **4.2.5 Data Treatment and Chemometric Analysis**

Figure 4-1 shows the general strategy proposed for chemometric analysis of time trend LC-HRMS data for target and industrial WWTP influent samples. The target data were used for the development of the two complementary strategies, as well as the validation script. Industrial wastewater samples served as proof of concept.

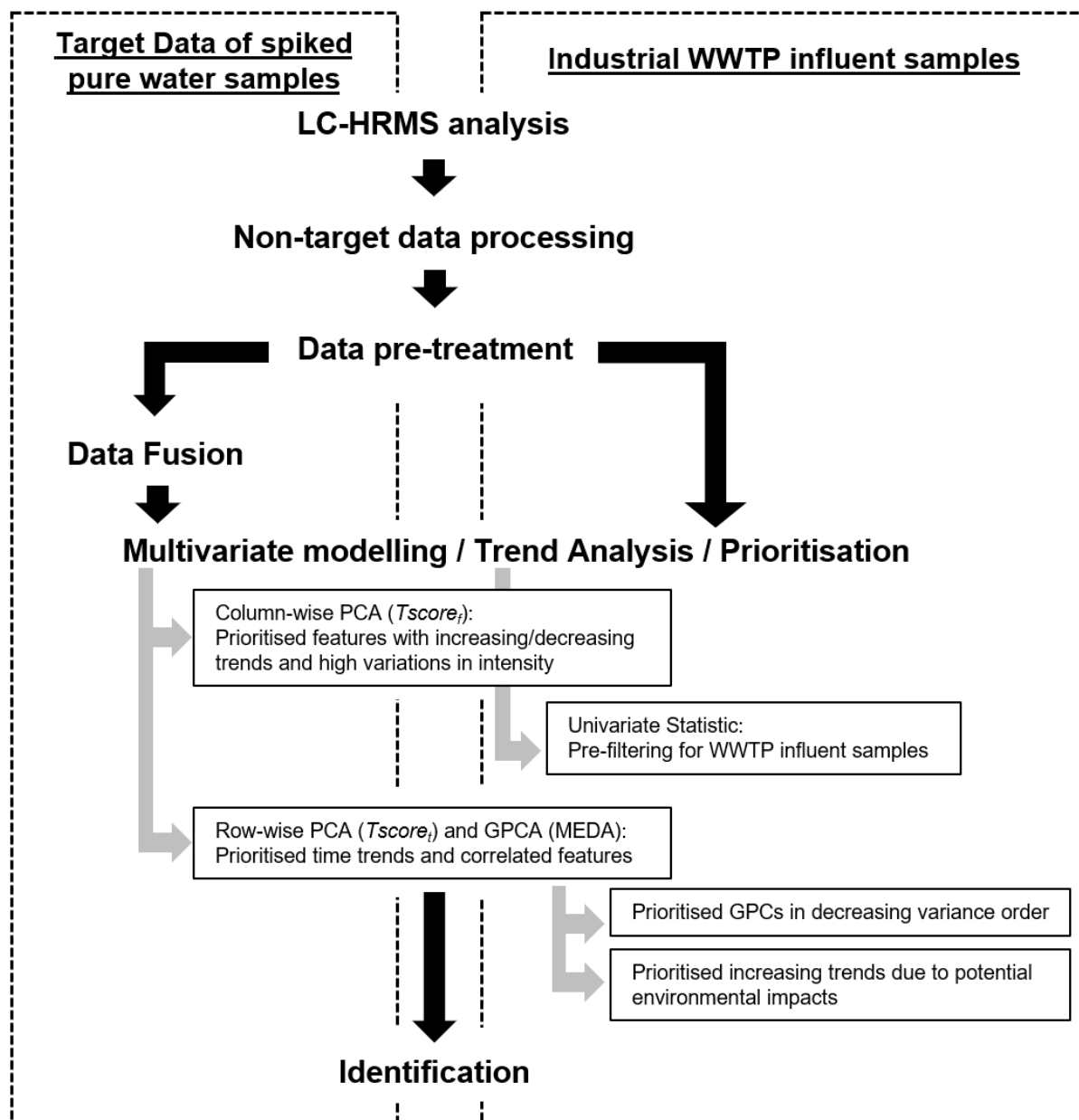


Figure 4-1 – Workflow of the data treatment, including LC-HRMS analysis (SCIEX OS), non-target data processing (MarkerView) and trend analysis (MATLAB). Before trend analysis, for target data, data fusion was performed. Subsequently, two complementary multivariate strategies of multivariate modelling were performed: 1. Column-wise PCA was performed on the data matrix having features in rows and time points in columns with  $Tscore_t$  ranking of features. 2. Row-wise PCA and group-wise PCA (GPCA) were performed on the matrix of data where rows are time points and columns are features. Here,  $Tscore_t$  ranked the time points and the groups of features were obtained. For real WWTP influent samples, the workflow of GPCA can further follow two possible routes, prioritising GPCs in decreasing variance order or prioritising increased trends due to environmental impacts. Additionally, for wastewater samples, a univariate statistic was applied as pre-filtering step. Trend analysis was followed by the identification of prioritised time trend relevant features (SCIEX OS, ChemSpider, MetFrag).

#### 4.2.5.1 Non-Target Data Processing Workflow

MarkerView (MV, version 1.3.1, SCIEX, Framingham, USA) was used for peak detection, peak alignment, normalisation of retention times and blank value subtraction. Import of raw LC-HRMS data files (.wiff) into MV occurred in two steps; the first step locates the peak in the



data and the second step performs the alignment and normalisation. For peak detection, the following criteria were used: retention time window: 1 – 26 min, background subtraction by the spectrum ten scans before the current one with a multiplication factor of 2.0, noise threshold: 1000 cps, minimum peak width: 30 ppm and minimum retention time peak width: 8 scans. Blank peaks were subtracted using overall exclusion. The alignment of multiple samples was performed using retention time (2.0%) and mass tolerances (3.0 ppm). Isotopes were removed and '*peaks with unknown status*' (where no automatic assignment of isotopes was possible) were kept. The average values of retention times of internal standards (ISTD, see Table-A 4-3) were used to correct the retention times of samples (mass tolerance of 2.0 ppm and retention time tolerance of 10%). Furthermore, zeros in the data matrix were replaced by the signal noise of the instrument. MV converted data to centroid data in the form of peak retention times and peak  $m/z$ . A data matrix was prepared that listed samples in rows and features in columns. The detected intensities of the features were the values of the data matrix. Subsequently, the generated data matrix of MV was imported into MATLAB® environment for further data processing.

#### 4.2.5.2 *The Fusion of Target Data for Constructing Data Matrix of Both Ionisation Modes*

For target data, the analysis was performed on fused negative and positive ionisation data, facing two data sources, with a common retention time index, but different  $m/z$  ratios and intensities. The data fusion enabled the discussion of all features in parallel. The total variance of the original data sources was different, due to the different performance of the two ionisation modes, which necessitates scaling of data for better comparability before analysing all features. Fusion is usually performed by using a scaling factor (the ratio of the first singular value of matrices). However, data scaling had no significant effect on the current results, due to the applied data pre-processing steps, which reduced both the skewness of the data and adapted variations of long-term data collection. For wastewater samples, positive and negative feature matrices were evaluated separately during the first route of data analysis, because of the amount of data and computational costs. The separate examination did not limit the compatibility to the validation data.

#### 4.2.5.3 *Multivariate Data Analysis Strategy*

Different processing steps for data analysis, such as pre-processing and multivariate trend analysis, were performed using an internal MATLAB® script (release R2017b, The MathWorks Inc., Natick, USA). GPCA was computed using the commands available in MEDA toolbox (Camacho et al., 2015).

#### 4.2.5.3.1 Row and Column-wise PCA for Exploratory Analysis

The first step of the proposed data analysis strategy consisted of PCA assessment of pre-treated data matrices and their transposed matrices with  $M$  observations (time points) and  $N$  features. Among several data pre-processing approaches, logarithmic transformation was selected. Using this transformation, skewness was reduced, giving normally distributed variables. Furthermore, row-wise normalisation followed by autoscaling was applied due to different magnitudes and scales of features and also wastewater samples over time. Row-wise modelling in this study stands for multivariate modelling of a typical matrix arrangement where rows are time and columns are features. Direct feature exploration and prioritisation were evaluated using column-wise PCA (or PCA of transposed initial data matrix) and calculation of  $Tscore_t$  values, which are a combination of the D-statistic and the Q-statistic values in a single selecting score (cf. 4.6.1.1, equation IV). Besides, row-wise PCA was applied to the data matrices determining  $Tscore_t$  indices, which were presented as the first evaluation score for time trend detections. A time point with high  $Tscore_t$  reflects the high contribution of some features for abnormal intensity obtained at that time.

#### 4.2.5.3.2 GPCA for Feature Prioritisation

As a complementary approach, GPCA was performed on the auto-scaled matrix of typical matrix arrangement where rows are time and columns are features (comparable to row-wise PCA). Using GPCA, which is a recent sparse variant of PCA, every component contains non-zero loadings for a single group of correlated features. GPCA consists of three parts: using MEDA approach (missing-data for exploratory data analysis) for computation of correlation maps, identification of the groups of associated variables using group identification algorithm (GIA) and fitting GPCA model (José Camacho et al., 2017). MEDA is used to find relationships among variables in the data. Thus, several groups of features can be detected and visualised, as has been shown recently (Jose Camacho et al., 2019). The most important GPCs are prioritised based on the amount of captured variance or the relevance of features included in each group considering the knowledge about the system or research interests. For wastewater trend detection, Spearman's rank correlation was used as a pre-filtration step for detecting increasing or decreasing trends. A more detailed description regarding the mentioned chemometrics methods can be found in the Appendix 4.6.1.

#### 4.2.5.4 Identification of Relevant Features

For identification, library search was performed. Proposed formulae were determined by the SCIEX OS software in 'Non-targeted Screening' workflow (Version 1.4.1, SCIEX GmbH, Darmstadt; Germany), the online database *ChemSpider* (Royal Society of Chemistry, 2017)

and the *High-Resolution Accurate Mass Libraries* (HRAM All-in-one v.1.1, contains spectra for 2231 compounds, SCIEX GmbH, Darmstadt; Germany). Furthermore, recorded MS/MS spectra were selected for further processing using the *in silico* fragmentation simulations of *MetFrag* (Aceña et al., 2015; Ruttkies et al., 2016). To classify tentatively identified features, the scheme of Schymanski et al. (2014) was used (Schymanski et al., 2014).

## 4.3 Results and Discussion

### 4.3.1 Multivariate Time Trend Analysis of Target Data

#### 4.3.1.1 Column-wise PCA for Prioritisation of Time Trends with High Variation in Intensity of Target Data

For finding features with increasing or decreasing trends, column-wise PCA (features were in rows) was performed investigating the inherent structure of the data. Groups were explored based on their principal components' indices (scores and loadings) identifying features most closely linked together. In Figure 4-2, the score plot is presented (see also Figure-A 4-1). The first two principal components (PC1 34.0% and PC2 25.5%) provided, in agreement to other studies, a general view on the feature variations of the target data set describing 65.5% of the data variation (Hetzl et al., 2015; Ponce-Robles et al., 2018; Siepak and Sojka, 2017). All relevant features, belonging to targets whose intensities followed increasing (e.g. azoxystrobin), decreasing (e.g. carbamazepine, triglyme and cyproconazole) or both trends (e.g. prothioconazole; the intensity of the feature first decreased followed by an increase in intensity) over time, were clearly separated from the bulk of non-relevant features (see Figure 4-2 a). The group of features having both trends lie between those features with increasing and decreasing intensity trends over time. Furthermore, for example, prothioconazole showing a feature in both positive and negative ionisation mode was grouped by PCA demonstrating the usefulness of data fusion. In other case, these features would be discussed separately complicate evaluation. In literature, the application of data fusion was already shown for the comparison of LC-MS and LC with capillary electrophoresis data, demonstrating that resolving low abundant metabolites was more efficient after fusion (Ortiz-Villanueva et al., 2017).

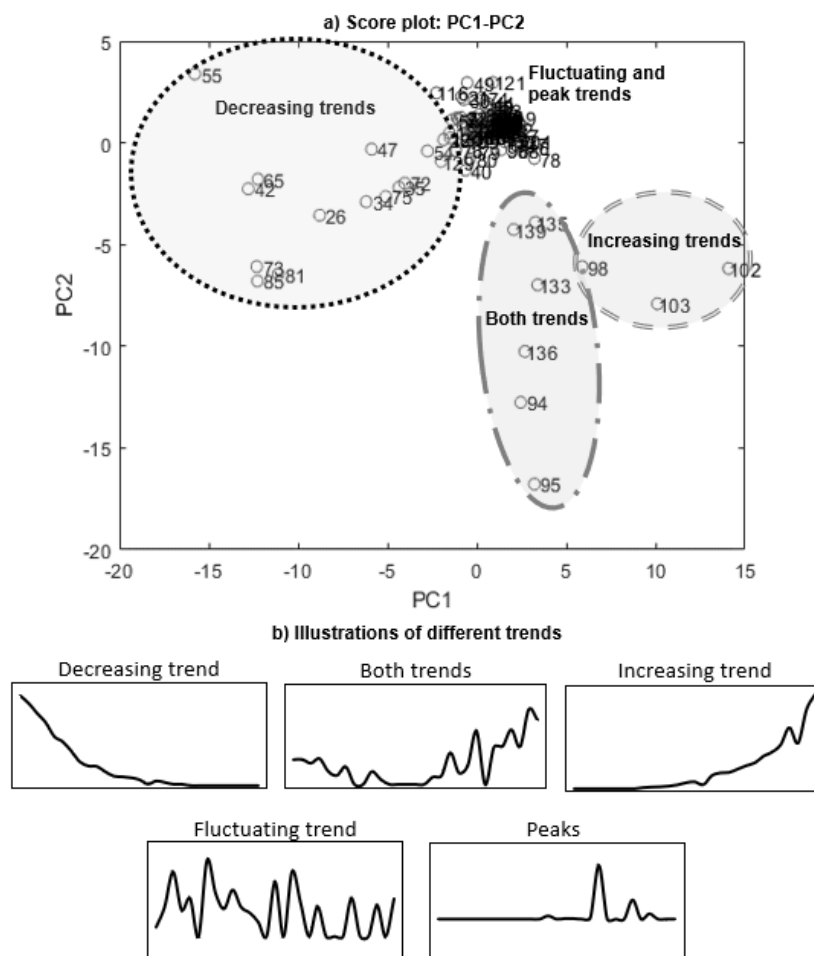


Figure 4-2 – Score plot obtained by PCA of target data matrix of (a) principal component 2 (PC2 25.47%) versus principal component 1 (PC1 39.99%). The ellipses in the figure highlight groups of trends exemplarily shown in (b).

$Tscore_f$  index, which is a combination of the D-statistic (Hotelling's  $T^2$ ) and Q-statistic (cf. also Figure-A 4-1), was calculated for every feature.  $Tscore_f$  index indicates significant intensity variations among daily composite samples over time illustrated as outliers or high leverage points from the bulk of uniform or low fluctuating trends in the PCA model. This was done because features of increasing/decreasing intensity trends with significant intensity variations, in agreement with other studies, seemed more relevant than those having only small variations in intensity (Albergamo et al., 2019; Alygizakis et al., 2019). In the first 30 prioritised features (total of 149, see section 3.2), three showed increasing-, ten had decreasing- and four showed both time trends, whereas 13 features were assigned to be non-relevant trends. Thus, 17 of 30 top-prioritised features corresponded to relevant ones of high interest because of high variation in intensity over time. The prioritised features were assigned to spiked target compounds; either to the main molecule adduct ( $[M+H]^+$ ,  $[M-H]^-$ ) or to one of their fragments. A detailed summary of assigned  $Tscore_f$  of top-prioritised features sorting back to target compounds according to 13 principal component model (with a total variance of 95.4%) can be found in Table-A 4-7. The rank of the  $Tscore_f$  result fits the spiked time trends of features

of highly fluctuated and continuously increasing/decreasing intensities over time (cf. Table-A 4-7 with spiked targets of Table-A 4-2). However, different fragments or adducts of the same compound (as a result of incomplete componentisation) may capture different ranks and also be prioritised by the  $Tscore_f$  index. For example, for carbamazepine three features were prioritised using  $Tscore_f$  index: feature with index 55 and  $m/z$  237.1016 ( $[M+H]^+$ ), feature with index 65 and  $m/z$  259.0841 ( $[M+Na]^+$ ) and feature with index 34 and  $m/z$  194.0962 (most intensive fragment). Consequently, this describes an advantage of using GPCA, since in this method, in contrast to PCA, loading vectors are used which contain non-zero values just for groups of variables (José Camacho et al., 2017). For LC-HRMS trend analysis, by analysing the data matrix with GPCA, all or most of the features belonging to one compound can be appropriately grouped. Thus, features related to different trends and compounds can be found. Features selected by  $Tscore_f$  ranking for decreasing/increasing trends match those identified in GPCA (cf. Table-A 4-7 and resulting groups of GPCA in Table-A 4-8).

#### 4.3.1.2 GPCA for Grouping Time Trends of Target Data

Performing GPCA on target data, an association map was built of all features as shown in Figure 4-3, which illustrates the MEDA plot for the target data set. The red-blue box on the lower right part of the plot (row numbers 110 – 149) belongs to straight-line signal intensities of 40 features without any trend.

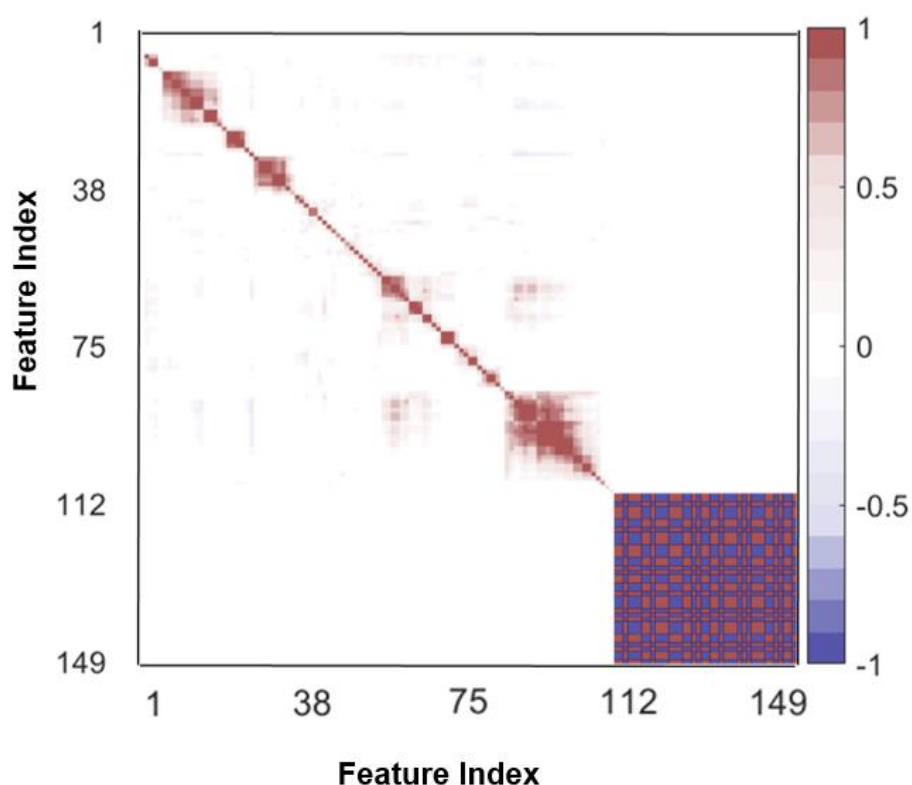


Figure 4-3 – Illustration of MEDA map for the set of target data, plotting feature index against each other. Colours in the plot reflect the level and direction (positive: red, or negative: blue) of the correlation between features. The red-blue box on the lower right part of the plot illustrates the features 110 to 149 with straight line trendless signals.

The data matrix was subjected to GIA for automatic identification of groups by defining the threshold  $\gamma$ . It was determined to capture relevant trends and to control sparsity of matrix (matrix in which most of the elements are zeros). The threshold was assigned by exploring the number and size of groups and preliminary experience regarding the target substances patterns. For example, if  $\gamma$  is set to 0.9, the group of features for azoxystrobin (increasing trend) including the features 98, 102 and 103 can be recovered, but at the same time, the whole trend related to features of prothioconazole (94, 95, 133, 135, 136 and 139; having both trends) was lost. Recovering this trend in a specific group, a  $\gamma$  value of 0.6 or less was needed. To prevent forming too many groups and generating sparse matrices, setting  $\gamma$  between 0.6 and 0.8 could be a good choice to capture the information in the squares with the GIA algorithm (José Camacho et al., 2017). Therefore, a medium level of 0.7 for  $\gamma$  was set as an optimum value (see Figure-A 4-4). Consequently, for the case of prothioconazole, the features 94, 95, 136 and 139 were categorised in GPC7 and features 133 and 135 were mainly grouped with features 98, 102 and 103 in GPC5. The two features of prothioconazole, which were assigned to a different group with the compromise threshold, showed a stronger expression of the increasing trend so that the assignment to azoxystrobin with an increasing trend became explainable. A value of  $\gamma = 0.7$  was, therefore, a good compromise, already used previously in

the literature, focussing on the basis of trends (José Camacho et al., 2017). The score and sparse loading vectors as well as groups of features for each trend, resulted by GPCA ( $\gamma = 0.7$ ), are shown in Figure-A 4-3 and Table-A 4-8, respectively.

#### 4.3.1.3 *Componentisation of Data Processing Method*

Data processing for NTS was performed by MV, removing isotopes. Features, where no isotope status could automatically be assigned, were kept. Therefore, more features than spiked targets were detected. For target data, this results in a total of 149 features in the pure water samples spiked with 19 target analytes in positive (111 features) and negative ionisation mode (38 features). Consequently, componentisation did not work properly with MV. However, using open-source workflows did also not guarantee complete grouping of features belonging to the same compound (Hohrenk et al., 2020). Componentisation is not included in all software packages, sometimes adduct and isotopic peaks are subtracted, the information is annotated or summarised into groups. However, by employing the MEDA map, which is a map of variables' correlation, followed by sparse matrix factorisation using GPCA, the feature similarities can be further explored. Therefore, multivariate modelling can be considered as a complementary technique to the MV incomplete componentisation. The described multivariate method, therefore, is not limited to prioritisation of continuously increasing/decreasing trends of intensity in time series but also enables a new componentisation pathway for NTS workflows. Using GPCA (see section 3.2.2), all or most of the features belonging to one compound can be appropriately grouped. (cf. Table-A 4-8).

### 4.3.2 **Application to Industrial WWTP Influent Wastewater Samples**

#### 4.3.2.1 *Pre-Filtering of Data Matrix using Univariate Statistic*

The screening experiment produced 2380 features in positive and 923 in negative ionisation mode aligned in 69 samples. These values reflect the removal of isotopes but may include adducts, multiply charged ions or in-source fragments as discussed above for the target data. Considering a large number of features, a wide range of temporal patterns and general fluctuating time trends of features, the classification of features into distinct groups was not as straightforward as for the target data. Exemplarily, the features of 2D score plot in Figure-A 4-6 (Column-wise PCA, PC2-PC1 space) located at different parts of the score plot, show related patterns. Besides, considering feature  $Tscore_i$  indices, when all types of trends were modelled with each other, a noticeable number of prioritised features were assigned to highly fluctuating trends, which were no trends of interest in the present study. This problem resulted in the detection of non-significant groups of several closely related TrOCs during time trend

survey. Thus, for wastewater samples, a pre-filtered data matrix, rather than full data, enhances the interpretation possibilities of derived models (cf. 4.6.3.2.1). Therefore, the number of features was reduced in an additional step resulting in relevant time trends of increasing/decreasing intensity. A Spearman rank's correlation with a threshold of  $|r| > 0.3$  (determined by target data, 4.6.3.2.1) was considered for prioritising and selecting those features whose intensity changed increasingly and decreasingly with time. However, this threshold should be evaluated for other studies in different matrices. For example, Chiaia-Hernández et al. (2017) used a threshold of  $|r| > 0.5$  pre-filtering LC-HRMS data time series for investigating contaminants of anthropogenic origin in two lakes in Central Europe (Chiaia-Hernández et al., 2017). By pre-filtering, in this study, 234 features were selected among 2380 of positive ionisation mode (9.8%) and 59 features were selected among 923 features of negative ionisation mode (6.4%). The results of applying the PCA on the pre-processed initial and pre-filtered data matrices, as 2D score and loading plots (a typical bi-plot visualisation) of first two principal components, are shown in Figure-A 4-7 for both ionisation modes.

#### 4.3.2.2 *Prioritisation of Relevant Features in WWTP Influent Samples*

Analysing features in wastewater time series, applying the validated chemometric methods of target data, resulted in prioritised features with increase/decrease in intensity over time. As was expected for LC-HRMS data sets, there was a high degree of correlations between features or mass fragments in wastewater samples. The MEDA map of pre-filtered data matrices discriminated the different groups of correlated features according to their increasing/decreasing intensity trends over time similarity degree which were visualised previous to MEDA by  $Tscore_i$  values (see Figure-A 4-8). The maps were obtained of the positive and negative ionisation data sets using 15 and 5 PCs (selected by *ckf* cross-validation, (Saccenti and Camacho, 2015)), respectively. MEDA map for positive data matrix is shown in Figure 4-4. As illustrated, features with increasing intensity trend over time comprised a small part of the MEDA map (Figure 4-4 b).



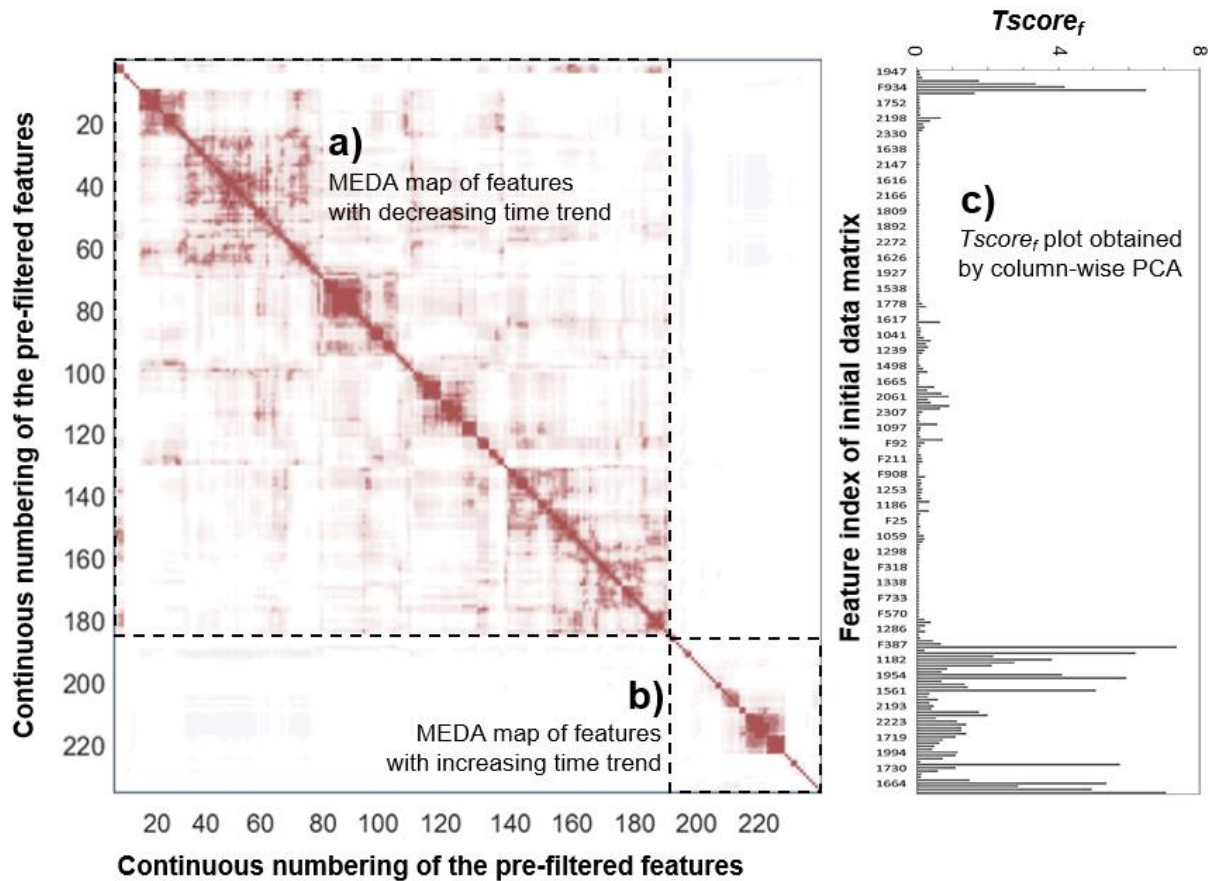


Figure 4-4 – MEDA map fitted with 15 components for pre-filtered WWTP influent data matrix for positive ionisation mode. (a) and (b) show the sub-maps related to decreasing and increasing trend features. (c)  $Tscore_f$  plot obtained by column-wise PCA modelling of pre-filtered WWTP influent data matrix in positive ionisation mode.

The workflow can continue by following different routes. The first route was in cases of both types of trends (= increasing and decreasing) being equally important and no information on a group of features was available. In this way, a global threshold  $\gamma = 0.7$  was set to capture the information of two squares (with both trends) using GIA algorithm. Thus, different groups of features were prioritised according to the GPCs in a decreasing variance order resulting in 25 and 8 GPCs as prioritised trends for positive and negative ionisation modes, respectively. A comparison was also made between MEDA map and  $Tscore_f$  indices (Figure 4-4 c) calculated on 234 features of the pre-filtered matrix of the positive ionisation mode. The first score vector with the highest variance had a decreasing trend. The second GPC contained most features responsible for increasing trends, especially at time point 48. Taking a look on time points within the pre-filtered time series,  $Tscore_f$  prioritised time points 2 to 7 and 48 to 60. The timeframes indicated the sudden presence of TrOCs in the WWTP influent samples, which could be a matter of concern. Therefore, the attention might focus on these particular timeframes to elucidate the time-related features. Thus, the developed strategies enabled additionally the prioritisation of sudden events by using  $Tscore_f$  index. In the case of time point

48, which belongs to sampling time in February, the high  $Tscore_i$  index and the grouping of time-related features in GPC2 mirrors a typical production cycle in the industry, starting production after December. In general, most of the GPCs belonged to decreasing trends confirming seasonal production of the studied industrial production processes. In the monitored time period consequently, some productions were reduced. Detailed results of GPCA of the first route are found in Figure-A 4-9, Figure-A 4-10, Figure-A 4-11, Table-A 4-9 and Table-A 4-10, including score plots and the features responsible for the increasing/decreasing trends as well as a detailed list of the first 30 prioritised features with  $m/z$  and retention time.

As the second route, in a non-target time trend study, the interest was in the examination of the increasing trends due to potential environmental impacts. To this end, a subset of relevant (increasing intensity trend over time) features was selected using the MEDA map of Figure 4-4 b either based on current related GIA states ( $\gamma = 0.7$ ) or based on updated GIA states (with different thresholds). In this study, this route was demonstrated for the positive and negative ionisation mode data subset (increasing trends) with updated threshold. To simplify the problem, the MEDA map obtained from the pre-filtered data matrix (size  $69 \times 44$ ) was subjected to GIA with  $\gamma = 0.6$  to capture the TrOCs with lower correlations. Afterwards, from this set of groups, six GPCs were extracted (see Figure 4-5) and the important components were further prioritised according to Spearman's rank correlation which revealed the strongest increase in intensity over time.

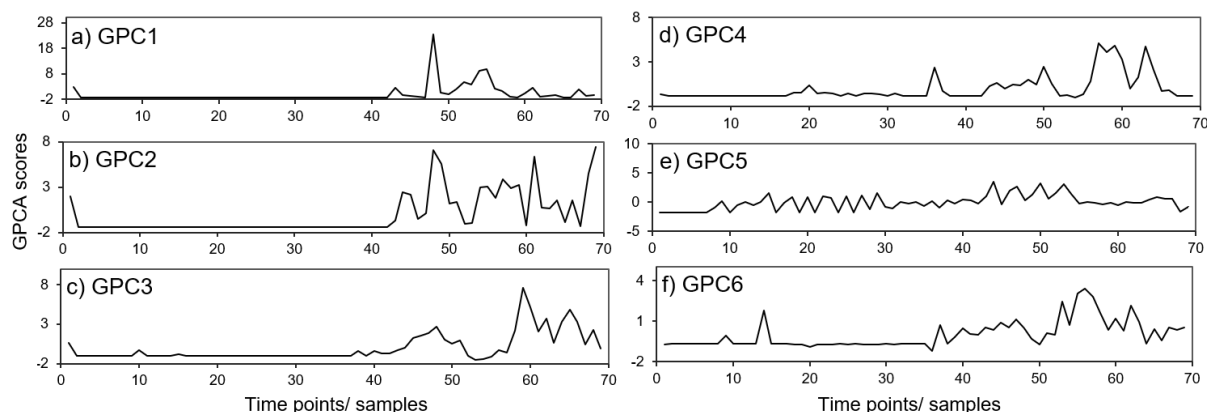


Figure 4-5 – Score vectors of six components from the GPCA model for the pre-filtered WWTP influent data matrix for positive ionisation mode. The prioritisation order is GPC6, GPC3, GPC1, GPC2, GPC5 and GPC4, respectively.

Consequently, the most relevant GPC was GPC6, followed by GPC3, GPC1, GPC2, GPC5 and GPC4. The recovered groups of correlated features in all GPCs were only from positive ionisation mode. Figure 4-5 and Table 4-1 present the temporal GPC scores and corresponding prioritised features.

Table 4-1 – GPCA results for feature prioritisation of WWTP influent samples containing the whole subset of continuously increasing trends of positive ionisation mode.

<b>GPC</b>	<b>Rank*</b>	<b>Features in groups</b>							
GPC1	3	F1883	F1994	F2110	F2031	F1937	F1808	F1719	F1474
		F1870	F1978	F2062	F2131	F2223	F2268	F1543	
GPC2	4	F2200	F2193	F2136	F2161	F1805	F1756		
GPC3	2	F1756	F1549	F1696	F1468	F1561			
GPC4	6	F1857	F1954	F1529					
GPC5	5	F1182	F188						
GPC6	1	F1529	F227						

\*Ranking of GPCs performed using Spearman's rank correlation values.

All in all, the prioritisation routes were proven capable of detecting intensity courses and compounds exhibiting increasing or decreasing trends over time. However, in long-term trend detection, false results can occur by matrix effects coming from different sample matrices. For example, the influent and effluent of a WWTP or even different surface waters have unique matrices. These different matrices can lead to ion suppression or enhancement, which in turn makes it appear that there are differences in concentration when there are not or vice versa (Nürenberg et al., 2015). Therefore, at least the behaviour of ISTD in different sample types should be analysed to detect those matrix effects. Unfortunately, a correction is only applicable to target compounds. For non-target compounds, the matrix effects cannot be compensated entirely (Nürenberg et al., 2015).

#### 4.3.2.3 Identification Experiment

The advanced prioritisation method could be further used to identify the most relevant contaminants present in WWTP influent samples from November 2018 to March 2019. For example, the 33 features of the second prioritisation route with relevant trends of increasing intensity over time could be subjected to further structural elucidation (analysis of increasing trend only, see Table 4-1). As proof of concept, the feature with *ID* 227 (*m/z* 100.0755, RT = 4.88 min), belonging to the most prioritised GPC6 having an increasing trend with high variation in intensity, was identified. The signal increased more than three times over the sampling period (max = 4.50E+03 cps) compared to average intensity (1.91E+03 cps, see Figure 4-6 c). Based on the isotope pattern of this compound, the molecular formula contains no chlorine, no bromine and no sulphur. In total, the following elements were chosen for the formulae finding: C<sub>0-30</sub>, H<sub>0-60</sub>, N<sub>0-10</sub>, O<sub>0-10</sub> and P<sub>0-5</sub>. Considering 5 ppm mass tolerance, one sum formula (as [M+H]<sup>+</sup>) was suggested: C<sub>5</sub>H<sub>9</sub>NO (1.6 ppm, SCIEX OS). Besides, *ChemSpider* was used as a platform to identify the elemental composition of the substance and resulted in more than 400 hits having the highest score (78.9%) for N-methylpyrrolidone. This suggested structure agreed with the short retention time in reversed-phase chromatography.

Furthermore, the measured spectra that were suspected to represent methylpyrrolidone were further compared to the *in silico* fragments of *MetFrag*, which explained the detected MS/MS peaks of the feature (cf. Figure 4-6 d). Therefore, the predominant profile of the feature with *ID 227* ( $m/z$  100.0752, RT = 4.88 min) resulted in the probable identification of the N-methylpyrrolidone (Schymanski et al., 2014). Subsequently, in positive ionisation mode, N-methylpyrrolidone was confirmed with the corresponding commercial standard, reaching the confidence level 1, as illustrated in Figure 4-6 (Schymanski et al., 2014).

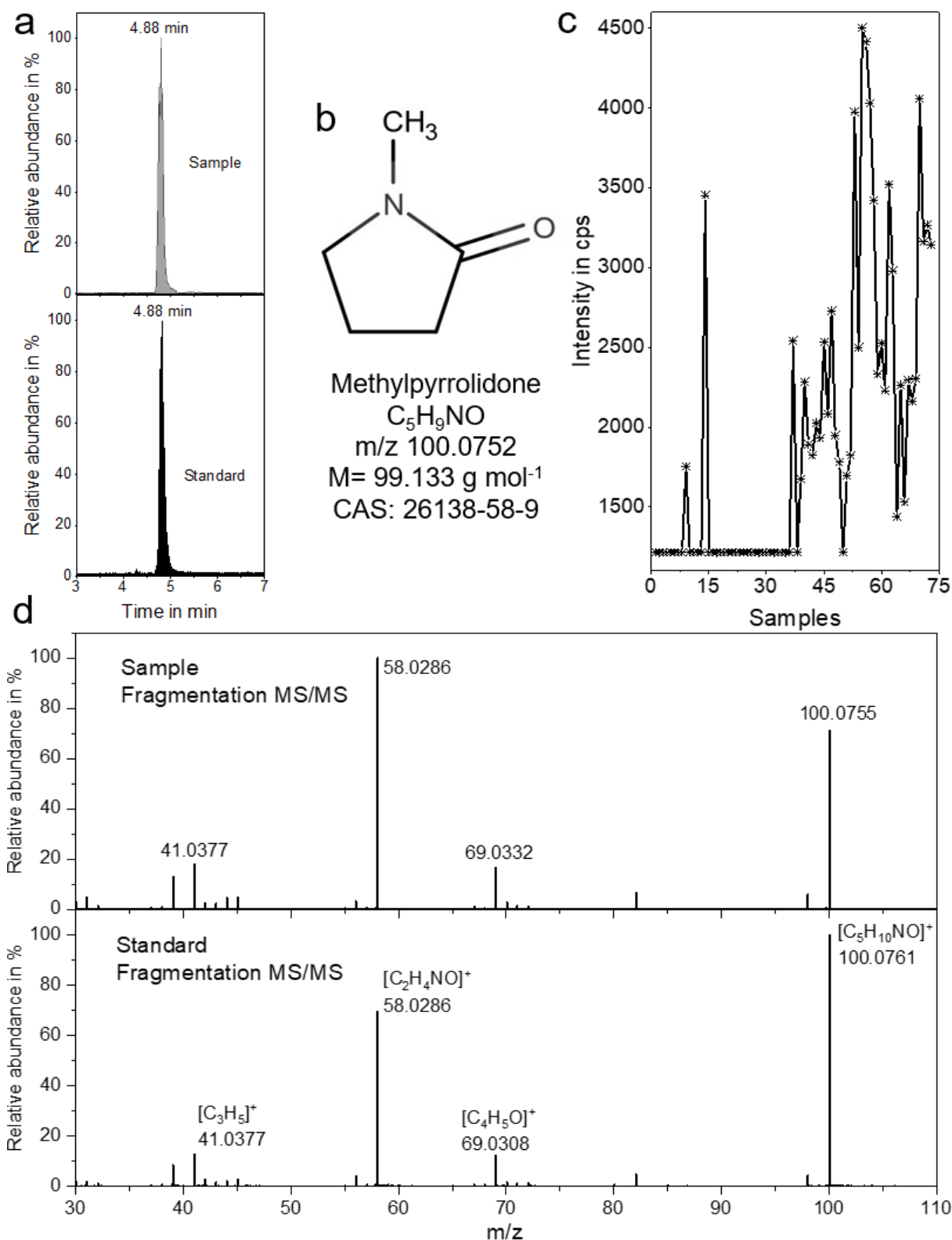


Figure 4-6 – Identification of the lactam *N*-methylpyrrolidone (Feature 227, [M+H]<sup>+</sup>). The extracted ion chromatograms (XIC) of sample and standard are shown in (a). In (b) the *N*-methylpyrrolidone structure are presented assigned by MarvinSketch (v.19.27, ChemAxon) and referenced by Sigma-Aldrich. The estimated wastewater intensity pattern over time is illustrated in (c). Tandem mass spectrometry spectra used for identification of *N*-methylpyrrolidone with the corresponding standard spectrum shown in (d).

Estimated concentrations of methylpyrrolidone were reported between 3.0 and 7 mg L<sup>-1</sup> for the last samples in the sampling period. There are several possible sources of this contaminant in industry. N-methylpyrrolidone is used as a solvent for polymers (acrylates, epoxies, polyurethanes, polyvinyl chloride and polyimides) and organic syntheses in the industry (Keshavarz et al., 2015; Nagpal and Rawat, 1981; Zeikus et al., 1999). However, during subsequent treatment in the WWTP, the compound was removed entirely, which was also shown in other studies (Cai et al., 2014).

#### 4.4 Conclusion

Development of new prioritisation methods capable of highlighting and identifying unknown compounds in wastewater samples is essential as NTS produces a large amount of data impossible to interpret without data reduction. Smart prioritisation strategies, combining the power of LC-HRMS with multivariate chemometric techniques focussing on relevant changes or properties of TrOCs, lead to a better understanding of water samples. This study demonstrated that in a non-target time trend approach, the combination of PCA with GPCA could be successfully used to obtain an adequate insight regarding complex industrial wastewater samples and prioritisation of groups of features corresponding to relevant time trends. Furthermore, followed by database mining, this strategy leads to the identification of prioritised unknown substances in industrial wastewater.

GPCA was applied in this study to overcome the disadvantages of PCA for exploratory time trend assessment. GPCA, in comparison to PCA, discriminated the unique variance of each variable from the shared variance among several variables, based on the idea of imposing sparsity in terms of groups of correlated variables. Therefore, it simplified extracted the hidden relationships between variables, which describes a necessity for interpretation of complex industrial wastewater time series. Furthermore, MEDA maps, as part of the GPCA, assigned correlations between features, whereas for standard covariance maps limitations regarding complex data could arise, as in MEDA, several spurious correlations are filtered out. Additionally, for LC-HRMS data, by analysing the data matrix with GPCA, all or most of the features belonging to one compound can be appropriately grouped. Thus, features related to different trends and compounds can be found. Therefore, the presented approach enabled the componentisation of features detected during NTS. In addition to componentisation, the grouping can also uncover TrOCs with correlating appearance/disappearance trends through investigating time series data. Thus, the developed chemometric strategies of this study could be applied in NTS workflows beyond wastewater time series. While this study limited the presentation of time trend analysis in industrial wastewater, the multivariate-assisted strategies

could also be efficiently employed for time trend exploration and extracting features in other scientific fields in which increasing/decreasing intensity trend in time play a role, e.g. in future biomarker searches. Besides, the evaluation of  $Tscore_t$  indices showed the applicability of prioritisation sudden events of TrOCs in contrast to prioritisation of increasing/decreasing intensity trend in time. Furthermore, combining a spatial correlation to time series could be used in further research to investigate, e.g. the elimination behaviour of WWTPs. Trends of TrOCs during the treatment procedure, over time and space, could be used for further prioritisation.

## 4.5 References

- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography–high-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 407, 6289–6299. <https://doi.org/10.1007/s00216-015-8852-6>
- Albergamo, V., Schollée, J.E., Schymanski, E.L., Helmus, R., Timmer, H., Hollender, J., de Voogt, P., 2019. Non-target screening reveals time trends of polar micropollutants in a riverbank filtration system. *Environ. Sci. Technol.* 53, 7584–7594. <https://doi.org/10.1021/acs.est.9b01750>
- Alygizakis, N.A., Gago-Ferrero, P., Hollender, J., Thomaidis, N.S., 2019. Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater. *J. Hazard. Mater.* 361, 19–29. <https://doi.org/10.1016/j.jhazmat.2018.08.073>
- Anliker, S., Loos, M., Comte, R., Ruff, M., Fenner, K., Singer, H., 2020. Assessing Emissions from Pharmaceutical Manufacturing Based on Temporal High-Resolution Mass Spectrometry Data. *Environ. Sci. Technol.* 54, 4110–4120. <https://doi.org/10.1021/acs.est.9b07085>
- Bader, T., Schulz, W., Lucke, T., 2016. Application of non-target analysis with LC-HRMS for the monitoring of raw and potable water: Strategy and results, in: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 2. ACS Symposium Series, Washington, DC.*, pp. 49–70. <https://doi.org/10.1021/bk-2016-1242.ch003>
- Brüggen, S., Schmitz, O.J., 2018. A New Concept for Regulatory Water Monitoring Via High-Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. *J. Anal. Test.* 2, 342–351. <https://doi.org/10.1007/s41664-018-0081-5>
- Camacho, J., García-Giménez, J.M., Fuentes-García, N.M., Maciá-Fernández, G., 2019. Multivariate Big Data Analysis for intrusion detection: 5 steps from the haystack to the needle. *Comput. Secur.* 87, 101603. <https://doi.org/https://doi.org/10.1016/j.cose.2019.101603>
- Camacho, J., Pérez-Villegas, A., Rodríguez-Gómez, R.A., Jiménez-Mañas, E., 2015. Multivariate Exploratory Data Analysis (MEDA) Toolbox for Matlab. *Chemom. Intell. Lab. Syst.* 143, 49–57. <https://doi.org/10.1016/j.chemolab.2015.02.016>
- Camacho, J., Rodríguez-Gómez, R.A., Saccenti, E., 2017. Group-Wise Principal Component Analysis for Exploratory Data Analysis. *J. Comput. Graph. Stat.* 26, 501–512. <https://doi.org/10.1080/10618600.2016.1265527>

- Camacho, J., Theron, R., Garcia-Gimenez, J.M., Macia-Fernandez, G., Garcia-Teodoro, P., 2019. Group-Wise Principal Component Analysis for Exploratory Intrusion Detection. *IEEE Access* 7, 113081–113093. <https://doi.org/10.1109/access.2019.2935154>
- Chiaia-Hernández, A.C., Günthardt, B.F., Frey, M.P., Hollender, J., 2017. Unravelling Contaminants in the Anthropocene Using Statistical Analysis of Liquid Chromatography–High-Resolution Mass Spectrometry Nontarget Screening Data Recorded in Lake Sediments. *Environ. Sci. Technol.* 51, 12547–12556. <https://doi.org/10.1021/acs.est.7b03357>
- Gowda, G.A.N., Alvarado, L.Z., Raftery, D., 2017. Chapter 5 - Metabolomics, in: *Nutrition in the Prevention and Treatment of Disease*. Elsevier Inc., pp. 103–122. <https://doi.org/10.1016/B978-0-12-802928-2.00005-9>
- Hedgespeth, M.L., Gibson, N., McCord, J., Strynar, M., Shea, D., Nichols, E.G., 2019. Suspect screening and prioritization of chemicals of concern (COCs) in a forest-water reuse system watershed. *Sci. Total Environ.* 694, 133378. <https://doi.org/10.1016/j.scitotenv.2019.07.184>
- Hetzel, T., Teutenberg, T., Schmidt, T.C., 2015. Selectivity screening and subsequent data evaluation strategies in liquid chromatography: the example of 12 antineoplastic drugs. *Anal. Bioanal. Chem.* 407, 8475–8485. <https://doi.org/10.1007/s00216-015-8994-6>
- Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.* 92, 1898–1907. <https://doi.org/https://doi.org/10.1021/acs.analchem.9b04095>
- Hohrenk, L.L., Vosough, M., Schmidt, T.C., 2019. Implementation of Chemometric Tools to Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes. *Anal. Chem.* 91, 9213–9220. <https://doi.org/10.1021/acs.analchem.9b01984>
- Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* 51, 11505–11512. <https://doi.org/10.1021/acs.est.7b02184>
- Huynh, N., Uddin, M., Minh, C.C., 2017. Chapter 10 - Data Analytics for Intermodal Freight Transportation Applications, in: Chowdhury, M., Apon, A., Dey, K. (Eds.), *Data Analytics for Intelligent Transportation Systems*. Elsevier, pp. 241–262. <https://doi.org/https://doi.org/10.1016/B978-0-12-809715-1.00010-9>
- Jalil, A., Rao, N.H., 2019. Chapter 8 - Time Series Analysis (Stationarity, Cointegration, and Causality), in: Özcan, B., Öztürk, I. (Eds.), *Environmental Kuznets Curve (EKC)*. Academic Press, pp. 85–99. <https://doi.org/https://doi.org/10.1016/B978-0-12-816797-7.00008-4>
- Kiefer, K., Müller, A., Singer, H., Hollender, J., 2019. New relevant pesticide transformation products in groundwater detected using target and suspect screening for agricultural and urban micropollutants with LC-HRMS. *Water Res.* 165, 114972. <https://doi.org/10.1016/j.watres.2019.114972>
- Lamchin, M., Lee, W.K., Jeon, S.W., Wang, S.W., Lim, C.H., Song, C., Sung, M., 2019. Corrigendum to “Mann-Kendall Monotonic Trend Test and Correlation Analysis using Spatio-temporal Dataset: the case of Asia using vegetation greenness and climate factors” (*MethodsX* (2018) 5 (803–807), (S2215016118301134)),



- (10.1016/j.mex.2018.07.006)). *MethodsX* 6, 1379–1383. <https://doi.org/10.1016/j.mex.2019.05.030>
- Mondal, A., Kundu, S., Mukhopadhyay, A., 2012. 70 Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of cuttack district, orissa. (Online) *An Online Int. J.* Available 2, 70–78.
- Nürenberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on liquid chromatography - high resolution mass spectrometry analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>
- Ortiz-Villanueva, E., Benavente, F., Piña, B., Sanz-Nebot, V., Tauler, R., Jaumot, J., 2017. Knowledge integration strategies for untargeted metabolomics based on MCR-ALS analysis of CE-MS and LC-MS data. *Anal. Chim. Acta* 978, 10–23. <https://doi.org/10.1016/j.aca.2017.04.049>
- Peter, K.T., Wu, C., Tian, Z., Kolodziej, E.P., 2019. Application of Non-Target High Resolution Mass Spectrometry Data to Quantitative Source Apportionment. *Environ. Sci. Technol.* 53, 12257–12268. <https://doi.org/10.1021/acs.est.9b04481>
- Plassmann, M.M., Tengstrand, E., Åberg, K.M., Benskin, J.P., 2016. Non-target time trend screening: a data reduction strategy for detecting emerging contaminants in biological samples. *Anal. Bioanal. Chem.* 408, 4203–4208. <https://doi.org/10.1007/s00216-016-9563-3>
- Ponce-Robles, L., Oller, I., Agüera, A., Trinidad-Lozano, M.J., Yuste, F.J., Malato, S., Perez-Estrada, L.A., 2018. Application of a multivariate analysis method for non-target screening detection of persistent transformation products during the cork boiling wastewater treatment. *Sci. Total Environ.* 633, 508–517. <https://doi.org/10.1016/j.scitotenv.2018.03.179>
- Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Sci. Total Environ.* 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Ruttkies, C., Schymanski, E.L., Williams, A.J., Krauss, M., 2019. Supporting non-target identification by adding hydrogen deuterium exchange MS / MS capabilities to MetFrag. *Anal. Bioanal. Chem.* 411, 4683–4700. <https://doi.org/https://doi.org/10.1007/s00216-019-01885-0>
- Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 1–16. <https://doi.org/10.1186/s13321-016-0115-9>
- Saccenti, E., Camacho, J., 2015. On the use of the observation-wise k-fold operation in PCA cross-validation. *J. Chemom.* 29, 467–478. <https://doi.org/10.1002/cem.2726>
- Samanipour, S., Reid, M.J., Thomas, K. V., 2017. Statistical Variable Selection: An Alternative Prioritization Strategy during the Nontarget Analysis of LC-HR-MS Data. *Anal. Chem.* 89, 5585–5591. <https://doi.org/10.1021/acs.analchem.7b00743>
- Schmidt, T.C., 2018. Recent trends in water analysis triggering future monitoring of organic micropollutants. *Anal. Bioanal. Chem.* 410, 3933–3941. <https://doi.org/10.1007/s00216-018-1015-9>
- Schollée, J.E., Schymanski, E.L., Hollender, J., 2016. Statistical Approaches for LC-HRMS Data to Characterize, Prioritize, and Identify Transformation Products from

Water Treatment Processes. ACS Symp. Ser. 1241, 45–65. <https://doi.org/10.1021/bk-2016-1241.ch004>

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. <https://doi.org/10.1021/es5002105>

Sharma, D., Kumar, B., Chand, S., 2019. A Trend Analysis of Machine Learning Research with Topic Models and Mann-Kendall Test. *Int. J. Intell. Syst. Appl.* 11, 70–82. <https://doi.org/10.5815/ijisa.2019.02.08>

Siepak, M., Sojka, M., 2017. Application of multivariate statistical approach to identify trace elements sources in surface waters: a case study of Kowalskie and Stare Miasto reservoirs, Poland. *Environ. Monit. Assess.* 189. <https://doi.org/10.1007/s10661-017-6089-x>

## 4.6 Chapter Appendix

### 4.6.1 Supplement Information of Chemometric Methods

#### 4.6.1.1 PCA for Exploratory Time Trend Assessment

Applying PCA to data matrices ( $\mathbf{X}$ ) with  $M$  observations (time points) and  $N$  features, the subspace of maximum variance in the  $N$ -dimensional feature space can be found. Principal components (PCs) which are linear transformations of original features are the eigenvectors of dispersion matrix of typically auto-scaled data. PCA model follows the expression:

$$\mathbf{X} = \mathbf{T} \times \mathbf{P}^T + \mathbf{E} \quad (\text{I})$$

where  $\mathbf{T}$  is the  $M \times A$  score matrix containing the projection of the objects onto the  $A$  number of PCs,  $\mathbf{P}$  is the  $N \times A$  loading matrix and  $\mathbf{E}$  is the  $M \times N$  matrix of residual. A pair of statistics are commonly used to detect abnormal behaviour in a system: the Hotelling's  $T^2$  statistic (D-statistic), which is computed from the scores and is a measure of the variation of each object within the PCA model and Q-statistic which is the residual between an object and its projection into the model. These statistics for each object are defined as follows:

$$D_m = \mathbf{t}_m \times \Lambda^{-1} \times \mathbf{t}_m^t \quad (\text{II})$$

$$Q_m = \mathbf{e}_m \times \mathbf{e}_m^t \quad (\text{III})$$

where  $\mathbf{t}_m$  is the score vector in the  $m^{\text{th}}$  row of  $\mathbf{T}$  in equation (I),  $\Lambda$  represents the covariance matrix of the scores and  $\mathbf{e}_m$  is the residual vector in the  $m^{\text{th}}$  row of  $\mathbf{E}$  in equation (I). Change in the behaviour of objects can be detected by inspecting the score scatter plots, where scores of some objects are different from the rest of the data. However, the information regarding the

objects and features (by PCA modelling of the transposed matrix) can be included in the Hotelling's  $T^2$  statistic since it is calculated using the score values. A scatter plot of Hotelling's  $T^2$  statistic versus Q-statistic is a preferable way of discerning any abnormal behaviour of objects based on exploratory PCA. Therefore, both statistics can be combined into a single selecting score, which is defined as  $Tscore$  (Jose Camacho et al., 2017) for each observation as follows:

$$Tscore_m = \alpha \times D_m / UCL^D + (1-\alpha) \times Q_m / UCL^Q \quad (IV)$$

where  $\alpha$  is a weighting factor for this combination and set to the total variance of the model (in percent).  $UCL^D$  and  $UCL^Q$  are the upper control limits for the D-statistic and Q-statistic (at 95% confidence level). This score was initially defined and implemented for intrusion detection in multivariate network monitoring studies (Jose Camacho et al., 2019; José Camacho et al., 2019). In the present work,  $Tscore$  was evaluated for time trend exploration and feature prioritisation while a typical LC-HRMS data matrix and its transpose subjected to PCA in the different modelling process.

#### 4.6.1.2 Group-wise PCA (GPCA)

GPCA is a sparse variant of the PCA where an adaptation of PCA to a simplified group-wise model can be justified by the data set (José Camacho et al., 2017). This method has been developed as one of the promising ways to overcome PCA disadvantages. The first disadvantage is that the unique variance for each variable cannot be discriminated from the shared variance among several variables, preventing the extraction of the hidden relationship between variables (Jolliffe, 2002). The second is that new latent variables (PCs) are usually a linear function of all variables, making the components difficult to interpret (Jolliffe et al., 2003). In GPCA, sparse solution is obtained in terms of groups of correlated variables. Here, every GPC contains non-zero loadings for a single group of correlated variables. So, the model and its interpretation will be considerably simplified. The first step of this approach is the computation of association map  $\mathbf{M}$  between variables using MEDA (Missing-data for Exploratory Data analysis (Camacho, 2011)) approach or any correlation matrix. Then,  $K$  groups of correlated/associated variables are identified using the so called group identification algorithm (GIA). For a correlation matrix of  $\mathbf{M}$  with the dimension of  $N \times N$  and elements between  $-1$  and  $1$ , a threshold  $\gamma$  (values between  $0$  to  $1$ ) can be defined in such a way that variables in  $\mathbf{M}$  matrix having the values larger than  $\gamma$  are selected for making a group ( $G$ ).  $\gamma$  should be optimised for each specific data set by inspecting the visualisation of MEDA and the output of the GIA, interactively. Following the groups' definitions, the GPCA algorithm first computes  $K$  candidate and sparse loading vectors associated with the variables in the  $k^{\text{th}}$

group. Then, according to the selected model, the loading with the largest explained variance is extracted from matrix **X**. further details regarding GPCA together with the algorithm is available in the literature (José Camacho et al., 2017).

## 4.6.2 Supplement of Experimental Section

### 4.6.2.1 Data sets, Target Compounds and Internal Standards

Table-A 4-1 – Table of data sets for data analysis method development and validation.

	<b>Validation data set</b>	<b>Authentic data set</b>
Samples	29 Milli-Q <sup>®</sup> samples spiked with 19 target analytes (see Table-A 4-2)	69 WWTP influent samples
Trends	Artificially produced trends over the 29 spiked samples	Authentic trends over five months: 2018-11 to 2019-03
Application	Method development	Verification of developed data analysis for authentic samples
Aim	Determination of temporal trends in the NTS	Validation of advanced data analysis

Table-A 4-2 – Spiked target analytes for training data set covering relevant (black) and non-relevant (grey) time trends (n= 29). All data are rounded data and refer to the LC-HRMS measurement. For the identification, a maximum mass error of ± 2 ppm and a maximum retention time error of ± 0.05 min were approved (\*Industrial chemical).

Target compounds	Sum formula	CAS No.	Doted conc. Range in µg/L	Mean conc. µg/L	Spiked Trend	Ionisation	Main Ion / Adduct	t <sub>R</sub> in min	m/z [ESI <sup>+</sup> ]	m/z [ESI <sup>-</sup> ]	Most intensive fragment
2-(Trifluoromethyl)benzamide	C <sub>8</sub> H <sub>6</sub> F <sub>3</sub> NO	360-64-5	0.200 - 30.0	0.5		positive	[M+H] <sup>+</sup>	6.73	190.0473	-	130.0294
Acifluorfen	C <sub>14</sub> H <sub>7</sub> ClF <sub>3</sub> NO <sub>5</sub>	50594-66-5	2.00 - 100	2		negative	[M-H] <sup>-</sup>	15.39	-	359.9889	315.9998
Azoxystrobin	C <sub>22</sub> H <sub>17</sub> N <sub>3</sub> O <sub>5</sub>	131860-33-8	0.200 - 200	20		positive	[M+H] <sup>+</sup>	14.51	404.1241	-	404.124
ABC 700*	C <sub>6</sub> H <sub>6</sub> F <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	Operating bottling	2.00 - 130	130		positive	[M+H] <sup>+</sup>	5.02	177.0468	-	177.0465
Carbamazepine	C <sub>15</sub> H <sub>12</sub> N <sub>2</sub> O	298-46-4	0.200 - 200	17		positive	[M+H] <sup>+</sup>	12.5	237.102	-	194.0963
Climbazole	C <sub>15</sub> H <sub>17</sub> ClN <sub>2</sub> O <sub>2</sub>	38083-17-9	0.200 - 17.0	0.5		positive	[M+H] <sup>+</sup>	11.74	293.105	-	69.0701

Table-A 4-2 continued.

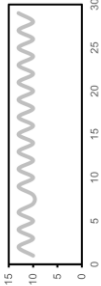
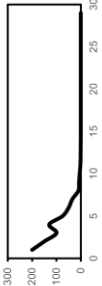
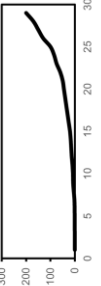
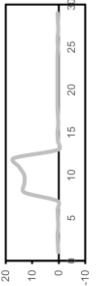
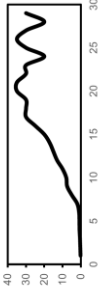
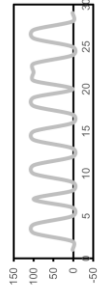
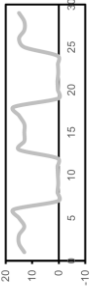
Clothianidin	<chem>C8H8ClN6O2S</chem>	2180880-92-5	10.0 - 13.0	10		positive	[M+H] <sup>+</sup>	7.82	250.0159	-	113.0166
Cyproconazol	<chem>C15H18ClN3O</chem>	94361-06-5	0.200 - 200	0.4		positive	[M+H] <sup>+</sup>	15.31	292.1214	-	70.0599
Diethylene glycol dimethyl ether/ Diglyme	<chem>C6H14O3</chem>	111-96-6	0.200 - 200	20		positive	[M+H] <sup>+</sup>	6.17	135.1012	-	59.0491
Diphenyl sulfone	<chem>C12H10O2S</chem>	127-63-9	0.200 - 17.0	0.5		positive	[M+NH4] <sup>+</sup>	11.81	236.0737	-	219.0472
Lactofen	<chem>C19H15ClF3NO7</chem>	77501-63-4	0.200 - 35.0	20		positive	[M+NH4] <sup>+</sup>	17.87	479.0828	-	200.2005
N,N-Diethyl-3-methylbenzamide/ DEET	<chem>C12H17NO</chem>	134-62-3	2.00 - 100	2		positive	[M+H] <sup>+</sup>	13.35	192.138	-	91.0538
N,N-Dimethyl-N'-phenylsulfamid/ DMSA	<chem>C8H12N2O2S</chem>	4710-17-2	0.200 - 17.0	13		positive	[M+H] <sup>+</sup>	9.21	201.0691	-	327.0075

Table-A 4-2 continued.

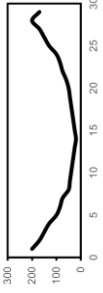
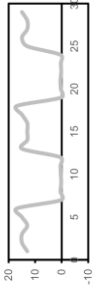
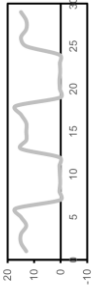
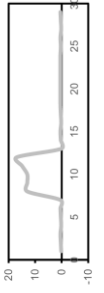
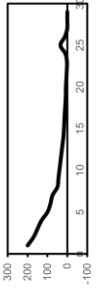
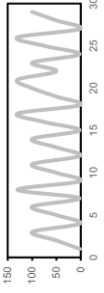
Prothioconazole	$C_{14}H_{15}Cl_2N_3OS$	178928-70-6	20.0 - 200	75		positive	[M+H] <sup>+</sup>	16.32	344.0388	-	98.9844
Tebuconazole	$C_{16}H_{22}ClN_3O$	107534-96-3	0.200 - 17.0	13		positive	[M+H] <sup>+</sup>	16.48	308.1526	-	125.0152
Tetraethyleneglycol dimethylether/ Tetraglyme	$C_{10}H_{22}O_5$	143-24-8	0.200 - 17.0	13		positive	[M+H] <sup>+</sup>	8.41	223.1539	-	59.0495
Thiacloprid	$C_{10}H_8ClN_4S$	111988-49-9	0.200 - 17.0	0.5		positive	[M+H] <sup>+</sup>	10.28	253.0343	-	126.0106
Triethyleneglycol dimethylether/ Triglyme	$C_8H_{18}O_4$	112-49-2	0.200 - 200	20		positive	[M+H] <sup>+</sup>	7.38	179.1277	-	40.9696
2-(Trifluoromethyl)benzenesulfonamide/ TBSA	$C_7H_6F_3NO_2S$	1869-24-5	2.00 - 130	50		negative	[M-H] <sup>-</sup>	7.06	-	223.9997	77.9658

Table-A 4-3– Table of used internal standards (ISTD). All data are rounded data and refer to the LC-HRMS measurement. For the identification, a maximum mass error of  $\pm 2$  ppm and a maximum retention time error of  $\pm 0.05$  min were approved. Fragment ions are listed in decreasing intensity from 1 to 5. Pos  $\triangleq$  positive ionisation mode, neg  $\triangleq$  negative ionisation mode. RT  $\triangleq$  Retention time.

Target compound	Azoxystrobin-D4	Bezafibrate-D6	Chloramphenicol-D5	Clothiazin-D3	m-Toluic acid diethylamide-D7 / DEET-D7	Diclofenac-D4	Diuron-D6	Meco-prop-D3	Mesotrione-D4	Metsulfuron-methyl-D3	Salicylic Acid-D4	Sulcotrione-D4	Tebuconazole-D6
ESI mode	positive	pos / neg	negative	positive	positive	pos / neg	pos / neg	negative	pos / neg	pos / neg	negative	pos / neg	positive
Main ion / adduct	[M+H] <sup>+</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M-H] <sup>-</sup>	[M+H] <sup>+</sup>	[M+H] <sup>+</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M-H] <sup>-</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M-H] <sup>-</sup>	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	[M+H] <sup>+</sup>
Sum formula	C <sub>22</sub> H <sub>13</sub> D <sub>3</sub> N <sub>3</sub> O <sub>5</sub>	C <sub>19</sub> H <sub>14</sub> D <sub>6</sub> ClNO <sub>4</sub>	C <sub>11</sub> H <sub>11</sub> D <sub>5</sub> C <sub>12</sub> N <sub>2</sub> O <sub>5</sub>	C <sub>8</sub> H <sub>8</sub> D <sub>3</sub> C <sub>12</sub> N <sub>2</sub> O <sub>2</sub> S	C <sub>12</sub> H <sub>11</sub> D <sub>7</sub> NO	C <sub>14</sub> H <sub>7</sub> D <sub>4</sub> NO <sub>2</sub> Cl <sub>2</sub>	C <sub>9</sub> H <sub>8</sub> D <sub>3</sub> Cl <sub>2</sub> N <sub>2</sub> O	C <sub>10</sub> D <sub>3</sub> H <sub>6</sub> ClO <sub>3</sub>	C <sub>14</sub> H <sub>12</sub> D <sub>4</sub> NO <sub>3</sub> S	C <sub>14</sub> H <sub>12</sub> D <sub>3</sub> N <sub>3</sub> O <sub>6</sub> S	C <sub>7</sub> H <sub>2</sub> D <sub>4</sub> O <sub>3</sub>	C <sub>14</sub> H <sub>9</sub> D <sub>4</sub> ClO <sub>3</sub> S	C <sub>16</sub> H <sub>10</sub> D <sub>6</sub> ClN <sub>2</sub> O
Molecular mass in g/mol	407.142	367.146	327.044	252.028	198.175	299.042	238.054	217.059	343.067	384.093	142.057	332.042	314.19
Expected mass in Da	408.149	368.153 / 366.139	328.036	253.035	199.182	300.049 / 298.035	239.062 / 237.047	216.051	344.074 / 342.059	385.1	141.05	333.050 / 331.035	314.19
Accurate mass in Da	408.1487	368.1527 / 366.1380	326.0362	253.0349	199.1822	300.0489 / 298.0343	239.0619 / 237.0473	216.0512	344.0737 / 342.0594	385.1003 / 383.0859	141.0494	333.0493 / 331.0348	314.1903
Mass error in mDa	-0.467	0.657471	-0.195	0.09	-0.036	0.538793	1.140351	-0.058	0.028 / 0.323	-2.279412	-0.08	1.01581	0.212
Mass Error in ppm	-1.1	0.666667	-0.6	0.4	-0.2	0.4 / -0.8	0.03 / -0.2	-0.3	0.1 / 0.9	-2.5	-0.6	1	0.7
RT in min	14.47	14.09	7.93	10.28	13.3	15.7	13.73	14.73	10.55	11.73	8.33	11.82	16.43
Concentration in µg L <sup>-1</sup>	0.025	25	25	25	1.25	25	25	25	25	25	25	0.25	1.25
1	377.128	138.9953 / 274.0642	157.066	126.011	126.094	218.0674 / 34.9691	78.0823 / 185.9522	216.033	-	170.0759 / 142.0813	97.0595	-	72.0537
2	349.134	121.0655 / 154.0067	34.9695	90.034	98.0985	254.0444 / 254.0446	52.1028 / 34.9693	34.9694	-	135.044	69.0648	-	279.232
3	-	322.1486 / 41.9989	45.9936	253.0316	58.9428	-	239.0608 / 237.0473	197.0487	-	199.0081	44.9981	-	95.0853
4	-	276.0801 / 90.0611	261.846	-	199.183	-	- / 41.9985	-	-	209.983	-	-	314.182
5	-	166.128	326.034	-	44.0122	-	- / 149.9749	-	-	-	-	-	-



#### 4.6.2.2 LC-HRMS Analysis

Table-A 4-4 – LC-HRMS acquisition method.

<b>Parameter</b>	<b>Setting</b>
LC system	Agilent 1290 Infinity
LC column	Raptor™ Ultra Aqueous C18 analytical column (100 mm x 2.1 mm, 3.0 µm particle size)
Mobile Phase A	0.1% formic acid in Milli-Q®
Mobile Phase B	0.1% formic acid in methanol
Gradient (%B)	0.5 min (0% B), 1.0 min (10% B), 20 min (90% B), 26 min (90% B), 32 min (0% B)
Flow rate	300 µL min <sup>-1</sup>
Temperature	55 °C
Sample injection volume	5 µL
MS system	x500R qTOF (SCIEX)
Ion source	Electrospray ion source Turbo V™
MS mass range	<i>m/z</i> 70 – 950
Ion source gas 1	50
Ion source gas 2	70
Curtain gas	40
Source temperature	450 °C
IonSpray voltage floating	5500 V / -4500 V
Declustering potential	80 V
Collision energy	7 V
MS/MS Experiment	10 data-dependent TOF-MS/MS scans
MS/MS mass range	<i>m/z</i> 30 – 700

#### 4.6.2.3 Analytes for Quality Control (QC)

Table-A 4-5 – Table of analytes and ISTD used for quality control (QC). All data are rounded data. Fragment ions are listed in decreasing intensity from 1 to 5. RT  $\pm$  Retention time, R  $\pm$  Recovery.

Target compounds	Sum formula	CAS No.	Main Ion / Adduct	RT in min	Expected Mass in Da	Found Mass in Da	Fragment ions in Da				
							1	2	3	4	5
1-(3-Sulphophenyl)-3-methyl-5-aminopyrazole	C <sub>10</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub> S	23646-86-8	[M-H] <sup>-</sup>	2.46	252.045	252.0446	79.9578	188.0831	107.0376	-	-
1,5-Diaminonaphthalene	C <sub>10</sub> H <sub>10</sub> N <sub>2</sub>	2243-62-1	[M+H] <sup>+</sup>	1.94	159.092	159.0915	115.0540	143.0740	159.0910	130.0650	89.0380
1,6-Naphthalenedisulfonic acid	C <sub>10</sub> H <sub>6</sub> O <sub>6</sub> S <sub>2</sub>	525-37-1	[M-H] <sup>-</sup>	2.03	286.969	286.9688	207.0123	143.0509	286.9693	79.9576	-
1,7-Naphthalenedisulfonic acid	C <sub>10</sub> H <sub>6</sub> O <sub>6</sub> S <sub>2</sub>	5724-16-3	[M-H] <sup>-</sup>	2.03	286.969	286.9687	207.0124	143.0509	286.9693	79.9576	-
2-(n-Ethylanilino)ethanol	C <sub>10</sub> H <sub>15</sub> NO	92-50-2	[M+H] <sup>+</sup>	2.68	166.123	166.1223	106.0650	120.0809	77.0387	166.1224	45.0333
2-(Trifluoromethyl)benzenesulfonamide/ TBSA	C <sub>7</sub> H <sub>6</sub> F <sub>3</sub> NO <sub>2</sub> S	1869-24-5	[M-H] <sup>-</sup>	4.96	224.000	223.9997	77.9656	160.0384	100.0192	224.0000	145.0282
2-(Trifluoromethyl)benzamide	C <sub>8</sub> H <sub>6</sub> F <sub>3</sub> NO	360-64-5	[M+H] <sup>+</sup>	5.32	190.047	190.0473	130.0300	102.0341	150.0358	170.0418	75.0234
2,2'-(4-Methylphenylimino)diethanol	C <sub>11</sub> H <sub>17</sub> NO <sub>2</sub>	430227	[M+H] <sup>+</sup>	3.25	196.133	196.1328	120.0806	134.0966	91.0543	196.1337	45.0336
2,4,6-Triamino-1,3,5-triazine/ Melamine	C <sub>3</sub> H <sub>6</sub> N <sub>6</sub>	108-78-1	[M+H] <sup>+</sup>	0.73	127.073	127.0724	85.0510	43.0289	68.0246	127.0727	110.0468
2,4,6-Trimethoxy-1,3,5-triazine	C <sub>6</sub> H <sub>9</sub> N <sub>3</sub> O <sub>3</sub>	-	[M+H] <sup>+</sup>	6.77	172.072	172.0714	72.0444	58.0288	83.0241	140.0461	-
2-Amino-4-methoxy-6-trifluoromethyl-1,3,5-triazine/ AMTT	C <sub>8</sub> H <sub>5</sub> F <sub>3</sub> N <sub>4</sub> O	1245970	[M+H] <sup>+</sup>	5.90	195.049	195.0487	110.0222	43.0291	57.0447	57.0447	68.9951
2-Chlorobenzoic acid	C <sub>7</sub> H <sub>5</sub> O <sub>2</sub> Cl	118-91-2	[M+H] <sup>+</sup>	6.72	157.005	157.0050	77.0387	111.0000	138.9951	51.0233	-
2-Naphthalenesulfonic acid	C <sub>10</sub> H <sub>6</sub> SO <sub>3</sub>	120-18-3	[M-H] <sup>-</sup>	5.29	207.012	207.0120	143.0505	207.0124	79.9575	115.0557	-
3,5-Dihydroxybenzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>	99-10-5	[M-H] <sup>-</sup>	2.66	153.093	153.0193	109.0299	41.0033	67.0190	65.0008	153.0244
3-Aminobenzenesulfonic acid	C <sub>6</sub> H <sub>7</sub> NO <sub>3</sub> S	121-47-1	[M-H] <sup>-</sup>	0.71	172.007	172.0072	79.9572	172.0078	108.0455	40.0193	-

Table-A 4-5 continued.

3-Hydroxy-2-nitrobenzoic acid	C <sub>7</sub> H <sub>5</sub> NO <sub>5</sub>	602-00-6	[M-H] <sup>-</sup>	3.44	182.009	182.0093	92.0266	45.9929	108.0235	138.0193	-
3-Nitrophthalic acid	C <sub>8</sub> H <sub>5</sub> NO <sub>6</sub>	603-11-2	[M-H] <sup>-</sup>	2.66	210.004	210.0041	45.9933	122.0249	163.9552	-	-
4-Aminobenzoic acid	C <sub>7</sub> H <sub>7</sub> NO <sub>2</sub>	150-13-0	[M+H] <sup>+</sup>	2.80	138.055	138.0549	77.0387	65.0386	92.0497	120.0449	51.0226
4-Chlor-2-cyano-5-(4-methylphenyl)imidazole/CCIM	C <sub>11</sub> H <sub>8</sub> ClN <sub>3</sub>	120118-14-1	[M+H] <sup>+</sup>	13.06	218.048	218.0478	183.0790	139.0310	130.0650	103.0550	118.0660
4-Chlorobenzoic acid	C <sub>7</sub> H <sub>5</sub> ClO <sub>2</sub>	74-11-3	[M-H] <sup>-</sup>	9.77	154.991	154.9904	34.9692	111.0009	44.9978	154.9964	-
4-Hydroxyphenylglycolic acid	C <sub>8</sub> H <sub>8</sub> O <sub>4</sub>	1198-84-1	[M-H] <sup>-</sup>	1.56	167.035	167.0350	-	-	-	-	-
4-Sulphophthalic acid	C <sub>8</sub> H <sub>6</sub> O <sub>7</sub> S	89-08-07	[M-H] <sup>-</sup>	0.97	244.976	244.9767	-	-	-	-	-
4-Toluenesulfonic acid	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> S	6192-52-5	[M-H] <sup>-</sup>	2.82	171.012	171.0121	79.9571	171.0120	107.0503	-	-
6-Chloropyridine-2-carboxylic acid	C <sub>6</sub> H <sub>4</sub> ClNO <sub>2</sub>	4684-94-0	[M+H] <sup>+</sup>	4.30	158.000	158.0002	111.9948	76.0183	139.9901	84.9840	75.0106
Acifluorfen	C <sub>14</sub> H <sub>7</sub> ClF <sub>3</sub> NO <sub>5</sub>	50594-66-5	[M-H] <sup>-</sup>	13.77	359.989	359.9890	194.9836	222.0303	286.0016	315.9991	45.9938
Azoxystrobin	C <sub>22</sub> H <sub>17</sub> N <sub>3</sub> O <sub>5</sub>	131860-33-8	[M+H] <sup>+</sup>	13.07	404.124	404.1234	372.0970	344.1026	329.0792	345.1056	330.0820
ABC 700	C <sub>6</sub> H <sub>8</sub> F <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	-	[M+H] <sup>+</sup>	3.56	177.047	177.1234	137.0345	42.0344	109.0399	177.0489	-
Benzamideoxime	C <sub>7</sub> H <sub>8</sub> N <sub>2</sub> O	613-92-3	[M+H] <sup>+</sup>	1.78	137.071	137.0709	77.0389	104.0500	51.0233	137.0712	119.0610
Benzenesulfonic acid	C <sub>6</sub> H <sub>6</sub> O <sub>3</sub> S	98-11-3	[M-H] <sup>-</sup>	1.79	156.996	156.9963	79.9576	93.0347	156.9967	-	-
Benzoic acid	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	65-85-0	[M-H] <sup>-</sup>	3.66	121.030	121.0296	77.0408	121.0266	92.0274	-	-
Carbamazepine	C <sub>15</sub> H <sub>12</sub> N <sub>2</sub> O	298-46-4	[M+H] <sup>+</sup>	10.84	237.102	237.1020	194.0958	179.0730	165.0700	152.0630	-
Chlorphenol (sum)	C <sub>6</sub> H <sub>5</sub> ClO	95-57-8	[M-H] <sup>-</sup>	7.75	126.996	126.9955	34.9694	91.0184	-	-	-
Ciprofloxacin	C <sub>17</sub> H <sub>18</sub> FN <sub>3</sub> O <sub>3</sub>	85721-33-1	[M+H] <sup>+</sup>	5.70	332.140	332.1400	314.1295	231.0570	332.1387	288.1498	98.9847
Climbazole	C <sub>15</sub> H <sub>17</sub> ClN <sub>2</sub> O <sub>2</sub>	38083-17-9	[M+H] <sup>+</sup>	10.14	293.105	293.1048	69.0699	197.0726	141.0104	225.0671	-
Clothianidin	C <sub>8</sub> H <sub>6</sub> ClN <sub>5</sub> O <sub>2</sub> S	2180880-92-5	[M+H] <sup>+</sup>	6.15	250.016	250.0161	131.9673	113.0168	169.0547	110.0722	70.9959
Cyproconazol	C <sub>15</sub> H <sub>18</sub> ClN <sub>3</sub> O	94361-06-5	[M+H] <sup>+</sup>	13.73	292.121	292.1209	70.0399	125.0152	138.9956	-	-

Table-A 4-5 continued.

Diethylene glycol dimethyl ether/ Diglyme	C <sub>6</sub> H <sub>14</sub> O <sub>3</sub>	111-96-6	[M+H] <sup>+</sup>	4.62	135.102	135.1015	59.0495	84.9597	116.9289	-	-
Diphenyl sulfone	C <sub>12</sub> H <sub>10</sub> O <sub>2</sub> S	127-63-9	[M+NH <sub>4</sub> ] <sup>+</sup>	10.15	236.074	236.0742	77.0390	219.0493	141.0004	-	-
Ditolyletherdisulfonic acid	C <sub>14</sub> H <sub>14</sub> O <sub>7</sub> S <sub>2</sub>	-	[M-H] <sup>-</sup>	3.97	357.011	357.0110	277.0550	185.9991	122.0378	79.9590	-
Diuron	C <sub>9</sub> H <sub>10</sub> Cl <sub>2</sub> N <sub>2</sub> O	330-54-1	[M+H] <sup>+</sup>	11.98	233.024	233.0241	72.0444	233.0238	159.9713	132.9614	187.9656
Fluopyram	C <sub>16</sub> H <sub>11</sub> ClF <sub>8</sub> N <sub>2</sub> O	658066-35-4	[M+H] <sup>+</sup>	13.82	397.054	397.0533	173.0215	208.0140	397.0544	1445.0260	130.0300
Imidacloprid	C <sub>9</sub> H <sub>10</sub> ClN <sub>5</sub> O <sub>2</sub>	138261-41-3	[M+H] <sup>+</sup>	6.70	256.060	256.0593	175.0978	209.0589	84.0541	128.0266	-
Ircadine	C <sub>12</sub> H <sub>23</sub> NO <sub>3</sub>	119515-38-7	[M+H] <sup>+</sup>	12.12	230.175	230.1749	130.1223	95.0856	112.1122	84.0808	156.1025
Lactofen	C <sub>19</sub> H <sub>15</sub> ClF <sub>3</sub> NO <sub>7</sub>	77501-63-4	[M+NH <sub>4</sub> ] <sup>+</sup>	16.53	479.080	479.0823	344.9975	222.9776	-	-	-
Maleic acid	C <sub>4</sub> H <sub>4</sub> O <sub>4</sub>	110-16-7	[M-H] <sup>-</sup>	1.10	115.004	115.0035	-	-	-	-	-
DEF 100*	C <sub>12</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub> S	-	[M+H] <sup>+</sup>	15.65	229.043	229.0429	110.9904	82.9951	-	-	-
Morpholine	C <sub>4</sub> H <sub>9</sub> NO	110-91-8	[M+H] <sup>+</sup>	0.64	88.076	88.0754	45.0333	70.0654	42.0338	55.0415	-
Moxifloxacin	C <sub>21</sub> H <sub>24</sub> FN <sub>3</sub> O <sub>4</sub>	151096-09-2	[M+H] <sup>+</sup>	7.70	402.182	402.1820	402.1824	384.1664	403.1892	364.1619	261.1060
N-(M-Tolyl)-diethanolamine	C <sub>11</sub> H <sub>17</sub> NO <sub>2</sub>	91-99-6	[M+H] <sup>+</sup>	3.72	196.133	169.1327	120.0807	134.0961	91.0542	45.0334	-
N,N-Diethyl-3-methylbenzamide/ DEET	C <sub>12</sub> H <sub>17</sub> NO	134-62-3	[M+H] <sup>+</sup>	11.67	192.138	192.1379	91.0538	119.0486	44.0130	-	-
N,N-Dimethyl-N'-phenylsulfamid/ DMSA	C <sub>8</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub> S	4710-17-2	[M+H] <sup>+</sup>	7.42	201.069	201.0693	92.0496	65.0389	137.1071	-	-
N-Allylthiourea	C <sub>4</sub> H <sub>8</sub> N <sub>2</sub> S	109-57-9	[M+H] <sup>+</sup>	2.08	117.048	117.0481	40.9699	84.9589	58.0662	-	-
N-Nitrosomorpholine	C <sub>4</sub> H <sub>8</sub> N <sub>2</sub> O <sub>2</sub>	59-89-2	[M+H] <sup>+</sup>	2.51	117.066	117.0657	45.0339	86.0599	57.0569	117.0661	91.0542
N-Phenylurea	C <sub>7</sub> H <sub>8</sub> N <sub>2</sub> O	64-10-8	[M+H] <sup>+</sup>	3.84	137.071	137.0708	77.0389	104.0496	51.0228	119.0604	137.0714
Phthalic acid	C <sub>8</sub> H <sub>6</sub> O <sub>4</sub>	88-99-3	[M-H] <sup>-</sup>	3.67	165.019	165.0193	77.0394	121.0296	96.9602	164.8977	147.8929

Table-A 4-5 continued.

Phthalic anhydride	C8H4O3	85-44-9	[M+H] <sup>+</sup>	15.98	149.023	149.0232	65.0388	39.0229	121.029 <sub>3</sub>	84.9602	149.0238
Phthalimide	C8H5NO2	85-41-6	[M+H] <sup>+</sup>	5.43	148.039	148.0392	130.0336	102.0341	75.0235	120.0446	51.0239
Picolinafen	C19H12F4N2O2	137641-05-5	[M+H] <sup>+</sup>	17.02	377.091	148.0392	102.0339	30.0331	75.0237	148.0400	-
Prothioconazole	C14H15Cl2N3OS	178928-70-6	[M+H] <sup>+</sup>	14.74	344.039	344.0387	98.9843	125.0152	189.047 <sub>1</sub>	326.0312	-
Py-ethylamine	C8H8ClF3N2	-	[M+H] <sup>+</sup>	4.51	225.04	225.0400	196.0136	208.0142	172.037 <sub>4</sub>	160.0371	30.0339
Salicylic acid	C7H6O3	69-72-7	[M-H] <sup>-</sup>	6.29	137.024	137.0243	93.0348	65.0399	-	-	-
Spiroxamine	C18H35NO2	118134-30-8	[M+H] <sup>+</sup>	12.00	298.274	298.2739	144.1379	100.1120	126.127 <sub>8</sub>	-	-
Tebuconazole	C16H22ClN3O	107534-96-3	[M+H] <sup>+</sup>	14.94	308.152	308.1524	70.0397	125.0155	151.031 <sub>3</sub>	165.0467	-
Tetraethyleneglycol dimethyl ether/ Tetraglyme	C10H22O5	143-24-8	[M+H] <sup>+</sup>	6.82	223.154	223.1540	59.0489	103.0755	207.033 <sub>1</sub>	191.0017	-
Thiacloprid	C10H9ClN4S	111988-49-9	[M+H] <sup>+</sup>	8.63	253.031	253.0305	126.0108	90.0340	-	-	-
Thiourea	CH4N2S	62-56-6	[M+H] <sup>+</sup>	0.71	77.017	77.0168	59.9901	43.0291	56.9638	-	-
Triethylamine	C6H15N	121-44-8	[M+H] <sup>+</sup>	1.88	102.128	102.1275	58.0649	102.1280	74.0968	30.0339	44.0496
Triethyleneglycol dimethyl ether/ Triglyme	C8H18O4	112-49-2	[M+H] <sup>+</sup>	5.80	179.128	179.1277	59.0493	103.0759	77.0399	133.0658	-
Trimethylcarboxylic acid	C12H13NO2	-	[M+H] <sup>+</sup>	8.80	204.102	204.1015	188.0706	174.0552	44.0810	160.1124	143.0728

### 4.6.3 Supplement of Results

#### 4.6.3.1 Results of Training Data

##### 4.6.3.1.1 Overview of Detection of ISTD in Training Data Set

Table-A 4-6 – List of labelled internal standards screened in positive and negative ionisation mode, detected ion/adduct, retention time, sample matches ( $n = 29$ ) and trend.

<b>Labelled-compound name</b>	<b>Main ion/ adduct</b>	<b>Retention time in min</b>	<b>Sample matches</b>
<i>Positive ionisation mode</i>			
Diuron-D6	[M+H] <sup>+</sup>	13.73	29
Bezafibrate-D6	[M+H] <sup>+</sup>	14.1	29
Diclofenac-D4	[M+H] <sup>+</sup>	15.71	29
Mesotrione-D4	[M+H] <sup>+</sup>	10.16	29
Metsulfuron-methyl-D3	[M+H] <sup>+</sup>	11.73	29
Sulcotrione-D4	[M+H] <sup>+</sup>	11.66	29
Clothianidin-D3	[M+H] <sup>+</sup>	7.79	29
Tebuconazol-D6	[M+H] <sup>+</sup>	16.45	29
DEET-D7	[M+H] <sup>+</sup>	13.3	29
Azoxystrobin-D4	[M+H] <sup>+</sup>	14.48	29
<i>Negative ionisation mode</i>			
Mecoprop-D3	[M-H] <sup>-</sup>	13.98	29
Bezafibrate-D6	[M-H] <sup>-</sup>	14.09	29
Diclofenac-D4	[M-H] <sup>-</sup>	15.7	29
Mesotrione-D4	[M-H] <sup>-</sup>	10.16	29
Metsulfuron-methyl-D3	[M-H] <sup>-</sup>	11.71	29
Sulcotrione-D4	[M-H] <sup>-</sup>	11.32	29
Diuron-D6	[M-H] <sup>-</sup>	13.73	29
Dicamba-D3	[M-H] <sup>-</sup>	9.73	29
Salicylsäure-D4	[M-H] <sup>-</sup>	8.28	29
Chloramphenicol-D5	[M-H] <sup>-</sup>	7.92	29
<i>Trends</i>			

4.6.3.1.2 Column-wise PCA

Figure S-1 (see SI 3.2) shows the scatter plot of Q-statistic versus D-statistic for exploration of features and identification of highly leveraged points (for example features 95, 40, 78 and 119 which exceed D critical limit and features 132, 84 and 47 which exceed Q-critical limit).

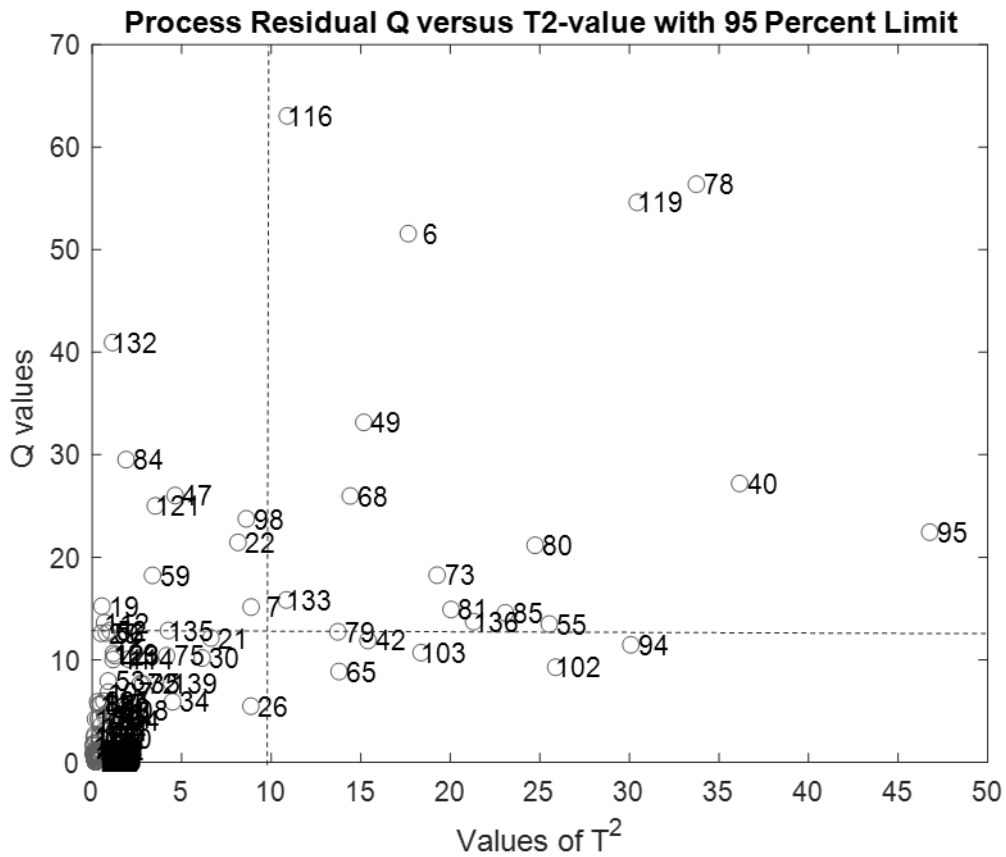


Figure-A 4-1 – Scatter plot of Q-values vs Hotelling's T<sup>2</sup> values obtained by PCA analysis of column-wise validation (target) data.

## Chapter 4 Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches

Table-A 4-7 – Detailed result of the 30 prioritised features (feature index, m/z and retention time) of column-wise PCA (the result of 13PCs) in combination with Tscore<sub>r</sub> index. ‘Both trends’ means first decreasing in intensity followed by increasing intensity changes over time. The spiked target analytes recovered are listed. Features assigned as unknowns did not fit spiked target compounds.

Index	m/z	Retention time in min	Tscore	Trend	Spiked target compounds
95	362.0307	15.98	5.1984	both trends	Prothioconazole
78	304.2608	20.63	4.1562	peaks-increasing	Unknown
40	201.0691	9.23	3.8173	fluctuating	N,N-Dimethyl-N'-phenylsulfamid/ DMSA
94	360.0334	15.98	3.2173	both trends	Prothioconazole
102	404.1238	14.54	2.4692	increasing	Azoxystrobin
55	237.1016	12.55	2.4297	decreasing	Carbamazepine
80	310.1498	16.52	2.2700	fluctuating	Tebuconazole
85	314.0638	15.36	2.2260	decreasing	Cyproconazol
6	137.0343	5.01	2.1450	fluctuating	ABC 700
73	292.1207	15.39	2.0146	decreasing	Cyproconazole
103	426.1063	14.55	1.8719	increasing	Azoxystrobin
81	312.0663	15.36	1.8696	decreasing	Cyproconazol
119	223.9998	7.21	1.6900	fluctuation	Clothianidin
68	282.2786	20.63	1.6878	peaks-increasing	Unknown
42	201.1093	7.47	1.6399	decreasing	Triethyleneglycoldimethylether/ Triglyme
136	358.0187	15.96	1.6147	both trends	Prothioconazole
49	214.1202	13.41	1.5438	fluctuating	N,N-Diethyl-3-methylbenzamide/ DEET
65	259.0841	12.56	1.4375	decreasing	Carbamazepine
79	308.1524	16.51	1.1794	fluctuating	Tebuconazole
22	177.0464	5.11	1.0005	fluctuating	ABC 700
98	372.0983	14.55	0.9687	increasing	Azoxystrobin
26	179.127	7.49	0.9497	decreasing	Triethyleneglycoldimethylether/ Triglyme
7	137.0343	5.12	0.9299	fluctuating	ABC 700
21	170.041	6.83	0.8850	peaks	2-(Trifluoromethyl)benzamide
133	342.0239	16.33	0.8412	both trends	Prothioconazole
30	190.0472	6.85	0.7922	peaks	2-(Trifluoromethyl)benzamide
116	175.0324	5.06	0.7726	fluctuating	2-(Trifluoromethyl)benzenesulfonamide/ TBSA
34	194.0962	12.55	0.5401	decreasing	Carbamazepine
47	210.1099	7.46	0.5236	peaks-decreasing	Triethyleneglycoldimethylether/ Triglyme
75	294.1181	15.38	0.5147	decreasing	Cyproconazol



#### 4.6.3.1.3 Row-wise PCA

Computing  $Tscore_t$  for time points ( $Tscore_t$ ) by row-wise PCA is the first evaluation step to show these time points which need further attention. Analysing target data, the time points (samples) 1, 3 and 20 of synthetic samples present higher scores which were justified considering the concentration of spiked compounds and their trend type. Also, an increasing trend is evident from time point 21 onward (cf. Figure S-2).

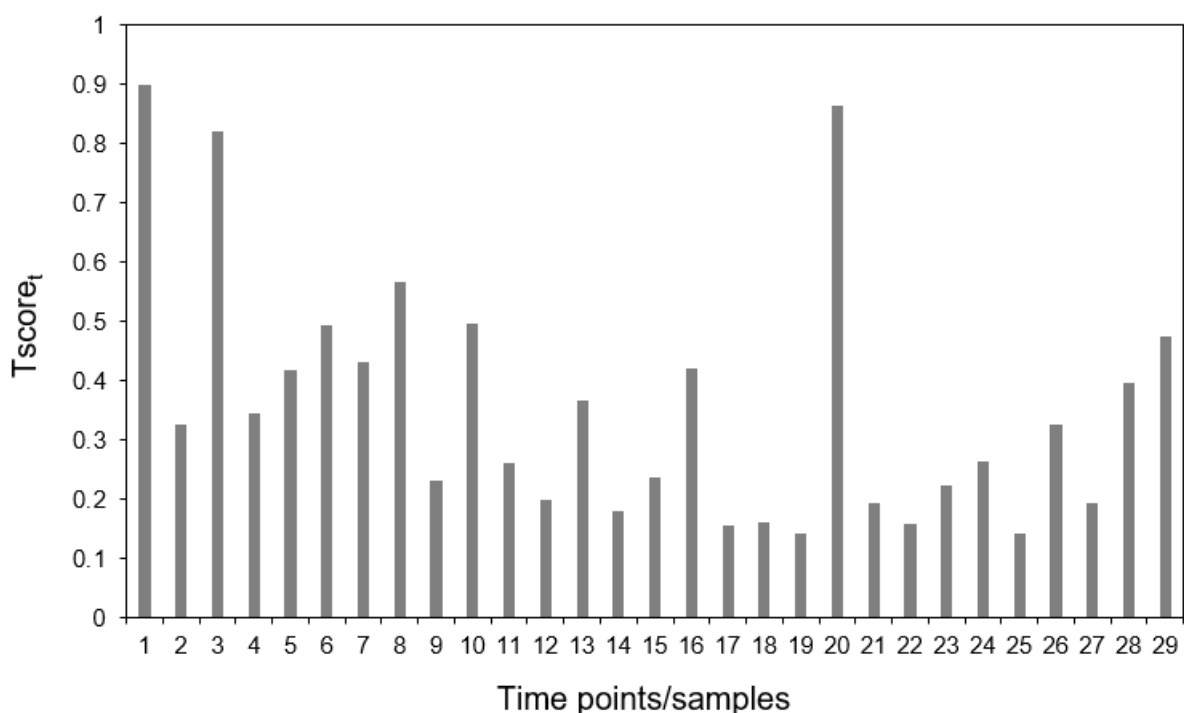


Figure-A 4-2 –  $Tscore_t$  plot obtained by row-wise PCA processing of validation data.

4.6.3.1.4 Results of GPCA

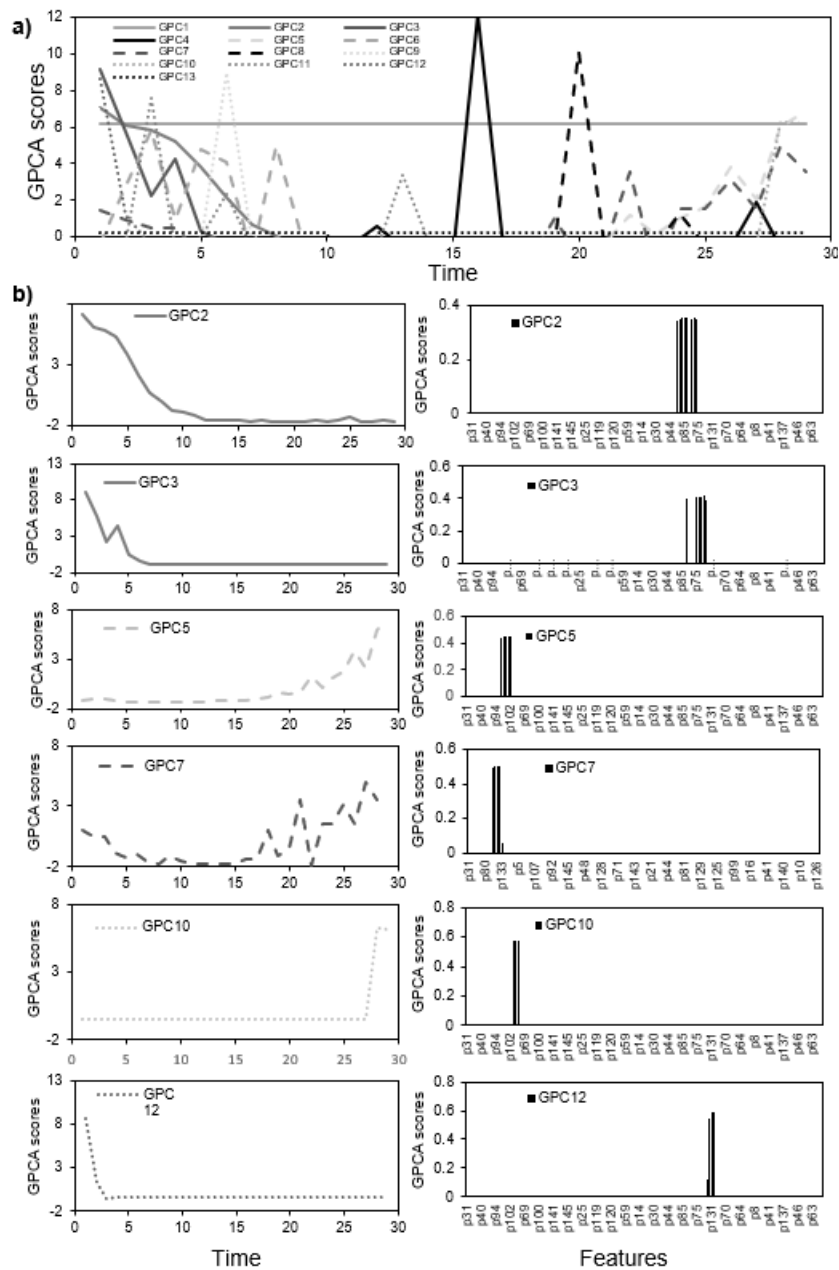


Figure-A 4-3 – Score vectors of first 13 components from the GPCA model for the validation (target) data set (a). Scores and loadings for the six prioritised decreasing/increasing trends using groups obtained from MEDA (b).

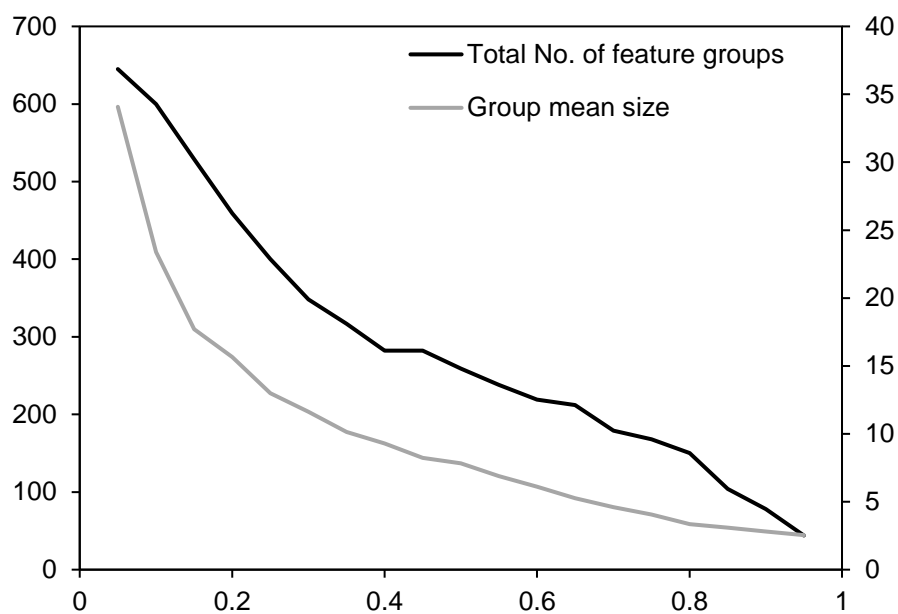


Figure-A 4-4 – Total number of feature groups and mean number of features per group as a function of selected threshold  $\gamma$  from the correlation map. The smaller the mean size of the group, the greater the sparsity. Nevertheless, by increasing the threshold value, some trends will be missed. So, initial experience about the association patterns of mass fragments is significant to select threshold value.

Table-A 4-8 – Groups of recorded features by GPCA modelling of validation (target) data.

Variance %	Cumulative variance	Trend	GPC	Feature indices	Target analyte
27.97	27.97	monotone	GPC1	2, 8, 9, 10, 13, 16, 17, 18, 20, 23, 24, 27, 33, 39, 41, 43, 45, 46, 57, 62, 63, 64, 86, 88, 89, 91, 93, 99, 105, 111, 115, 118, 124, 126, 127, 134, 137, 138, 140, 144	Unknown
5.328	33.30	decrease	GPC2	26, 34, 42, 55, 65, 73, 81, 85	Carbamazepine
3.994	37.29	decrease	GPC3	35, 72, 75, 76, 129, 130	Cyproconazole
3.970	41.26	single peak	GPC4	92, 113, 141, 142, 148, 149	Unknown
3.341	44.60	increase	GPC5	98, 102, 103, 133, 135	Azoxystrobin
3.277	47.88	fluct-decrease	GPC6	12, 28, 47, 54, 120	Carbamazepine
2.714	50.60	both trends	GPC7	94, 95, 133, 136, 139	Prothioconazole
2.716	53.31	single peak	GPC8	68, 69, 78, 108	Unknown
2.098	55.41	single peak	GPC9	3, 14, 97	Thiacloprid
2.068	57.48	increase	GPC10	5, 90, 96	Lactofen
2.052	59.53	peaks	GPC11	52, 56, 59	Tetraethyleneglycol dimethyl ether/ Tetraglyme
1.982	61.51	decrease	GPC12	11, 106, 130, 131	Unknown
1.399	62.91	single peaks	GPC13	19, 29	2-(Trifluoromethyl)benzamide

#### 4.6.3.2 Results of Wastewater Samples

##### 4.6.3.2.1 Pre-Filtering of Raw Data Matrix

For pre-filtering of wastewater samples data matrices two univariate statistics were analysed: the Spearman rank's correlation and the Mann-Kendall test (MK, which was initially used for climate factors analysis (Lamchin et al., 2019) and rainfall trend analysis (Mondal et al., 2012) or in the machine learning research (Sharma et al., 2019)). In these tests, the trend of intensities has been calculated for each feature individually. For the MK test, features were categorised to increasing/decreasing trends, which had a calculated p-value smaller than the significance level of 0.05 (Sharma et al., 2019). 75% of determined time trends assigned by MK confirmed the spiked time patterns, which results in 25% false negative MK determined time trends. Features which were not detected by MK had monotonic trends, showed both increasing and decreasing trend or having significant fluctuation while falling/increasing (cf. Figure S-5). The false negative results comprised, among others, the prioritised features by *Tscore* and GPCA. However, MK Test represents both increasing and decreasing trend in the intensity of features over time, although not much significant, as also determined by Modal et al. (Mondal et al., 2012). In comparison, the true and false detected trends using Spearman's rank correlation test were 94% and 6%, respectively. In other studies, the statistical performance from both statistical tests, MK and Spearman's rank correlation, were more consistent with each other (Ahmad et al., 2015; Shadmani et al., 2012; Yue et al., 2002). In addition, inspecting the recovered decreasing/increasing GPCs showed that almost all false non-significant trends (2 out of 24 features) did have a correlation coefficient of  $|r| < 0.3$ . Therefore, a threshold of  $|r| > 0.3$  (with a false negative rate of zero) was considered as a first screening step using Spearman rank's correlation for filtering the real data before performing multivariate processing.

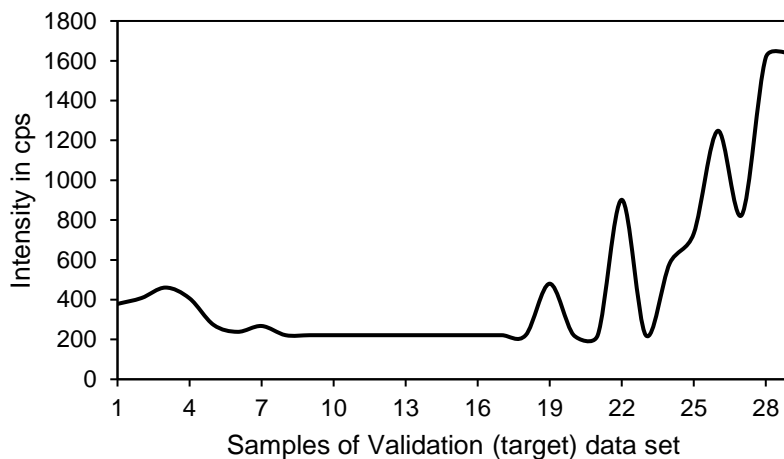


Figure-A 4-5 – Example of MK failed monotonic time trend test. Patterns with fluctuation in increasing/decreasing part of the trend cause failure in the MK test. The example shows the feature with ID 133 in positive ionisation mode. The  $p$ -value is 0.069, indicating no trend, according to MK ( $p > 0.05$ ).

Figure S-6 (see SI 4.1) shows a 2D score plot of features in PC2-PC1 space. Here, for example, features 1315 and 1410, 230 and 1198, 1870 and 1978 and 2223 located at different parts of the score plot, show a fluctuating pattern, peak pattern and an increasing trend ( $\rho = 0.68$ ), respectively.

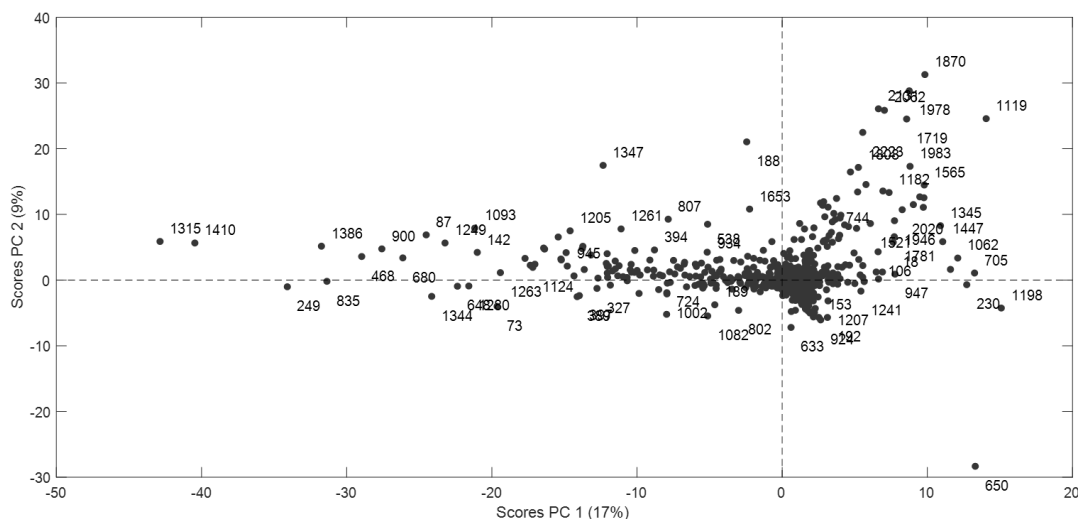


Figure-A 4-6 – 2D Score plot of the first two components derived from a PCA model based on LC-HRMS data matrix of influent WWTP sample in positive ionisation mode. Column-wise PCA model of pre-processed initial data matrix of positive ionisation mode, used 25 components to explain 80% of the total variance in the data.

The results of applying the PCA on the pre-processed initial and prefiltered data matrices, as 2D score and loading plots (a typical bi-plot visualisation) of first two principal components, are shown in Figure S-7 in both ionisation modes.

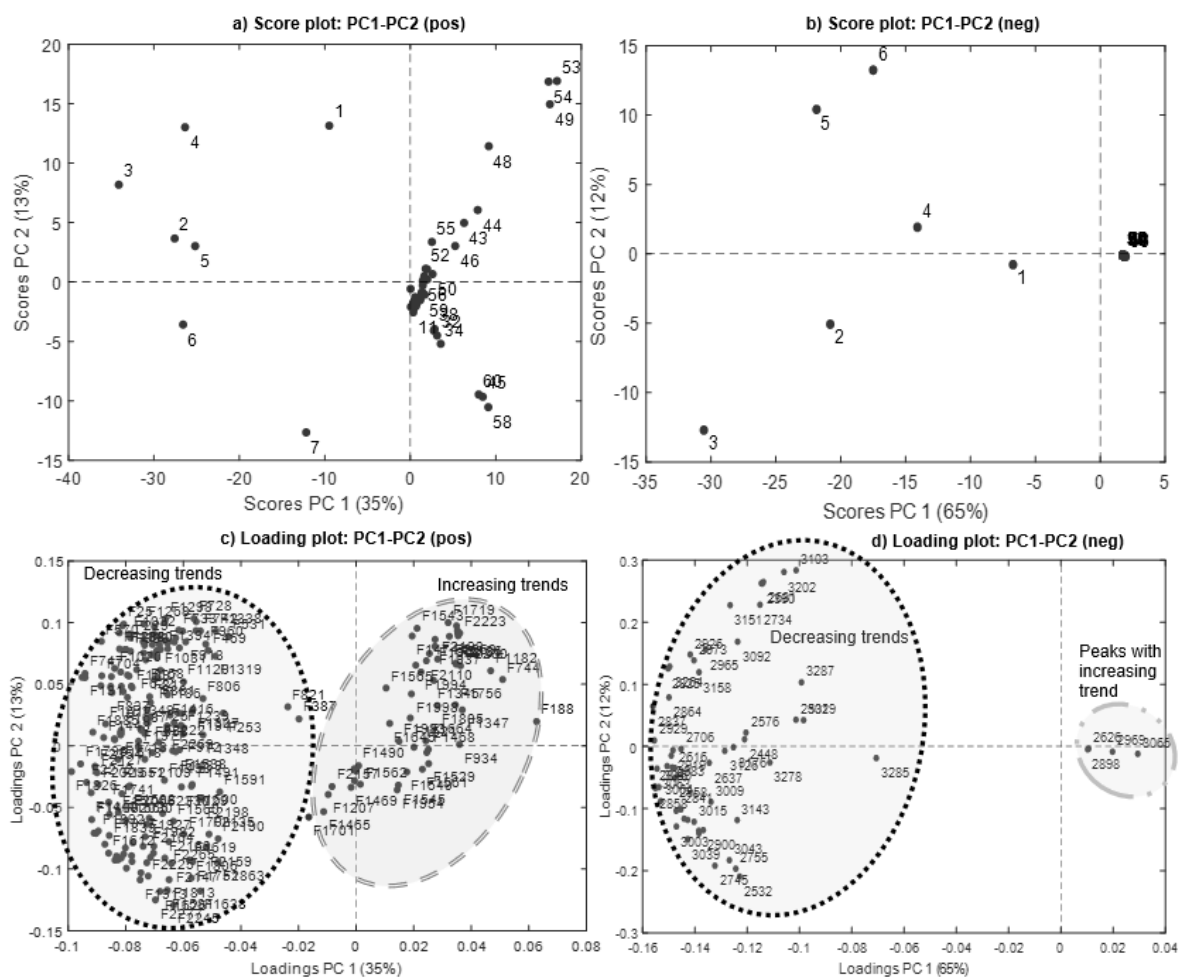


Figure-A 4-7 – 2D score (a & b) and loading (c & d) plots obtained by PCA modelling of pre-filtered WWTP influent data matrices containing increasing/decreasing intensity patterns for positive and negative ionisation mode, respectively.

#### 4.6.3.2.2 Results of Prioritising Samples in Time Series by Row-wise PCA

While a general overview regarding increasing and decreasing time trends in loading plots can be perfectly visualised, the definitive conclusion about the relationship between variables and time points (especially for increasing trends) is a hard task for complex data like industrial wastewaters. As an alternative visualisation of the problem, the evolution of  $Tscore_t$  (validated for target data, see SI 3.3) indices with time points before and after the feature selection step is presented in Figure S-8. For the initial data set and without filtering, the important time points can be detected of calculated  $Tscore_t$  values. Here, the initial time points and time points 48 – 59 have higher score values than the others. The result of Figure S-8 focused the attention on the particular timeframe or even specific sampling times because the sudden presence of TrOCs in the environment could be a matter of concern. Therefore, data pre-filtering was performed to prioritise the time Tscore points 2 to 7 and 48 to 60.

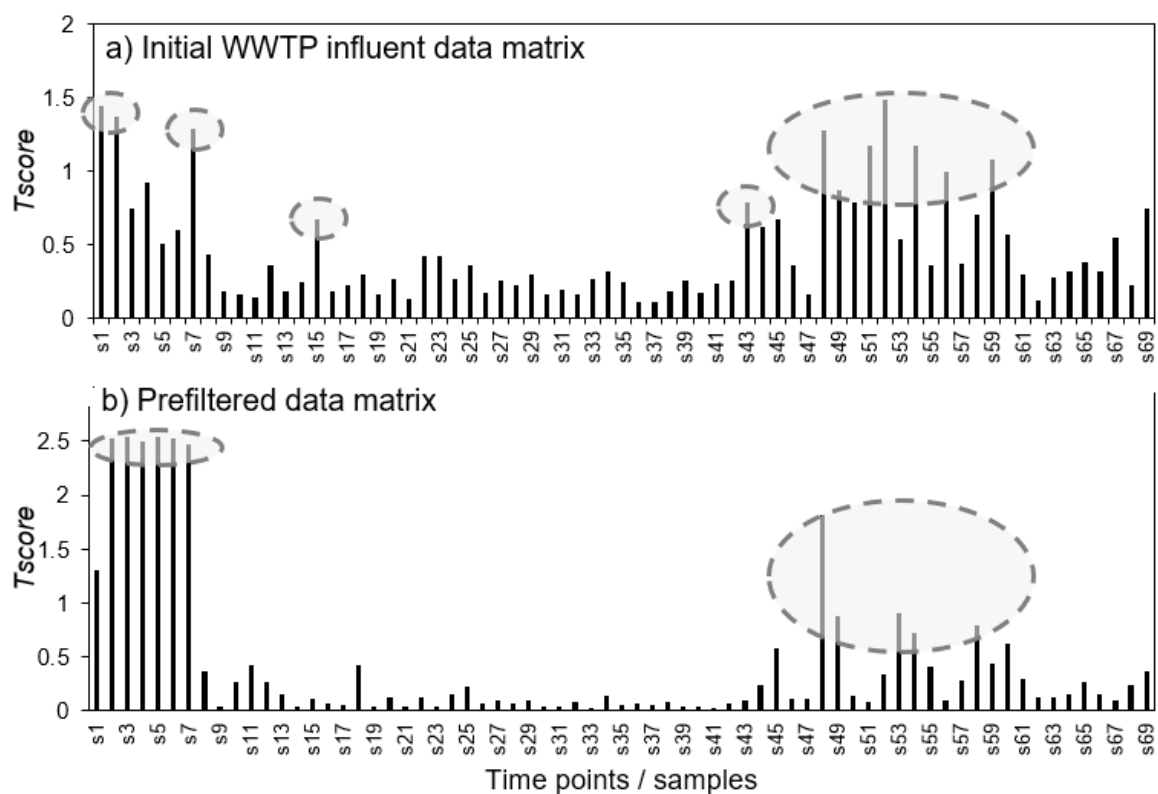


Figure-A 4-8 – Tscore values for PCA modelling of (a) initial WWTP influent data matrix (69\*2380) in positive ionisation mode using 15 PCs with a total variance of 81.1% and (b) prefiltered data matrix (69\*234) based on 10 PCs with a total variance of 87.3%. The most relevant time points regarding high score values are indicated by dashed circles.



4.6.3.2.3 GPCA Results and Prioritised Features Responsible for Trend Phenomena using First Prioritisation Route

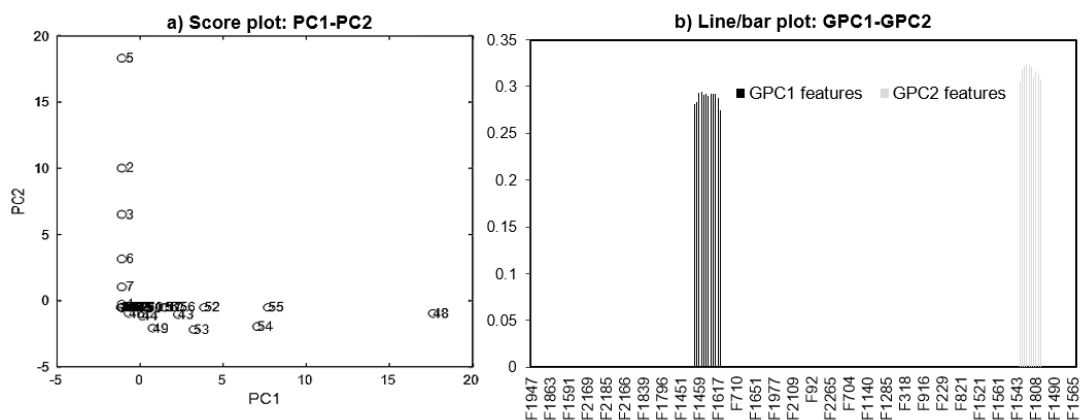


Figure-A 4-9 – 2D score plot of first 2 PCs in GPCA of pre-filtered WWTP influent data in positive ionisation mode (a) and the corresponding loading (line/bar) plot (b).

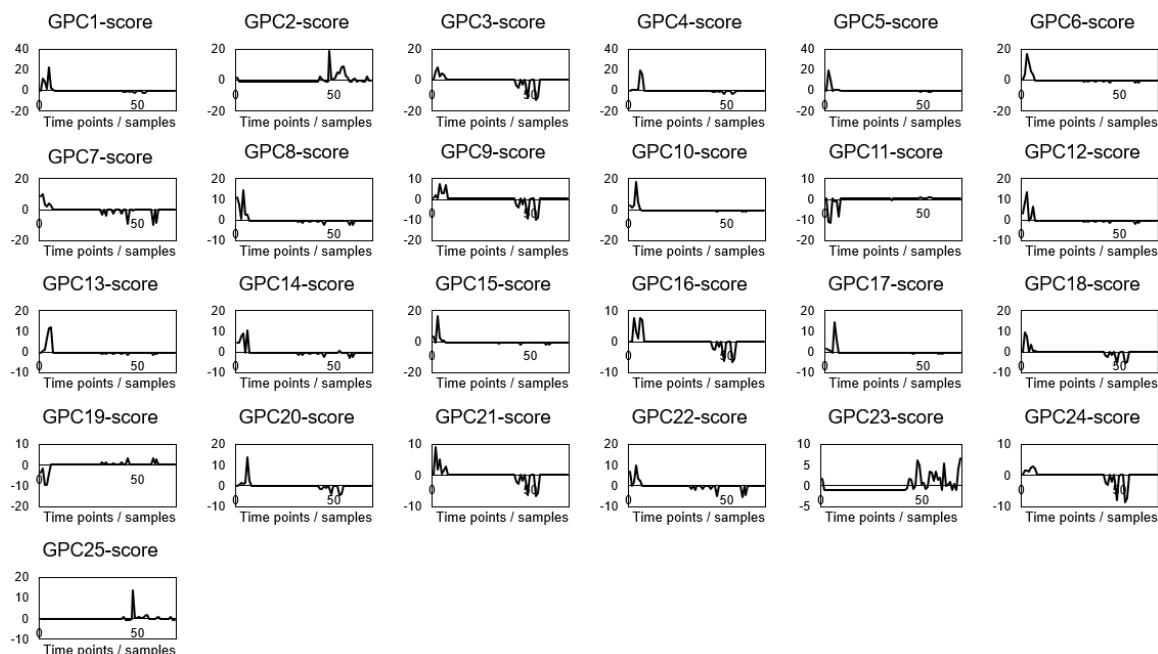


Figure-A 4-10 – Results of GPCA of positive ionisation mode. For example, the first score vector with the highest variance has a decreasing trend. The second GPC contains the most important features responsible for increasing trends, especially at time point/sample s48.

## Chapter 4 Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches

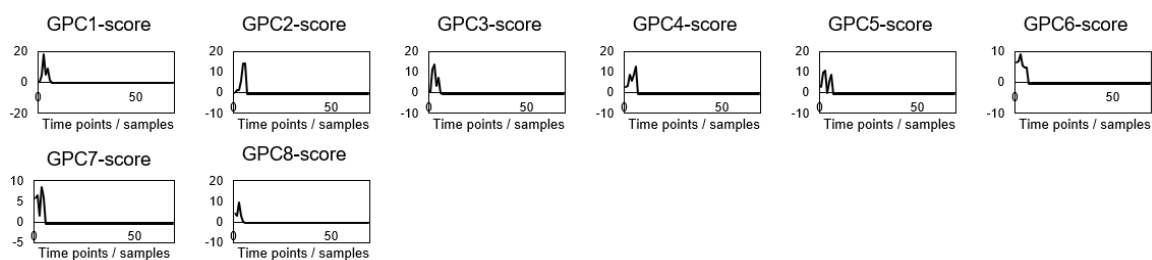


Figure-A 4-11 – Results of GPCA of negative ionisation mode. All GPCs belonged to decreasing trends.

Table-A 4-9 – Detailed list of features prioritised by GPCA. The features printed in thick are these ones following the relevant trend of continuously increasing/decreasing intensity. In total, 130 relevant features were prioritised.

Prioritised features of WWTP influent samples in positive ionisation mode												
<b>GPC1</b>	<b>F1457</b>	<b>F1459</b>	<b>F1618</b>	<b>F1687</b>	<b>F1778</b>	<b>F1686</b>	<b>F1713</b>	<b>F1714</b>	<b>F1617</b>	<b>F1717</b>	<b>F1538</b>	<b>F924</b>
<b>GPC2</b>	<b>F2062</b>	<b>F1978</b>	<b>F2131</b>	<b>F1870</b>	<b>F2223</b>	<b>F1808</b>	<b>F1937</b>	<b>F1719</b>	<b>F2031</b>	<b>F2268</b>	-	-
<b>GPC3</b>	<b>F1612</b>	<b>F2121</b>	<b>F2145</b>	<b>F1935</b>	<b>F2243</b>	<b>F2166</b>	<b>F1616</b>	<b>F1747</b>	<b>F1809</b>	<b>F1581</b>	-	-
GPC4	F2135	F2159	F1729	F1619	F1600	F2190	F1752	F1491	F1591	F2198	-	-
GPC5	F2269	F2307	F2359	F2109	F153	F725	F1307	F2222	-	-	-	-
GPC6	F747	F191	F1253	F1206	F1197	F1288	F622	F1380	-	-	-	-
<b>GPC7</b>	<b>F733</b>	<b>F913</b>	<b>F1232</b>	<b>F742</b>	<b>F916</b>	<b>F1429</b>	<b>F751</b>	<b>F728</b>	-	-	-	-
<b>GPC8</b>	<b>F1286</b>	<b>F1214</b>	<b>F959</b>	<b>F960</b>	<b>F1381</b>	<b>F227</b>	<b>F1529</b>	-	-	-	-	-
<b>GPC9</b>	<b>F2225</b>	<b>F2155</b>	<b>F1716</b>	<b>F2147</b>	<b>F2185</b>	<b>F2245</b>	<b>F2169</b>	<b>F2243</b>	-	-	-	-
GPC10	F1059	F894	F469	F1285	F1319	F1947	-	-	-	-	-	-
GPC11	F2359	F2222	F2130	F1722	F2061	F1977	F1873	-	-	-	-	-
<b>GPC12</b>	<b>F935</b>	<b>F1240</b>	<b>F1349</b>	<b>F1097</b>	<b>F192</b>	-	-	-	-	-	-	-
GPC13	F858	F362	F1041	F710	F837	-	-	-	-	-	-	-
<b>GPC14</b>	<b>F1020</b>	<b>F398</b>	<b>F1006</b>	<b>F92</b>	<b>F570</b>	<b>F1032</b>	-	-	-	-	-	-
GPC15	F1120	F806	F1186	F1140	F1245	F1206	F1197	-	-	-	-	-
<b>GPC16</b>	<b>F1896</b>	<b>F1917</b>	<b>F1913</b>	<b>F1488</b>	<b>F1574</b>	-	-	-	-	-	-	-
GPC17	F939	F1239	F1348	F1416	F1713	-	-	-	-	-	-	-
<b>GPC18</b>	<b>F1839</b>	<b>F2329</b>	<b>F1440</b>	<b>F2030</b>	<b>F2109</b>	-	-	-	-	-	-	-
GPC19	F1839	F2329	F1440	F2030	F2109	-	-	-	-	-	-	-
GPC20	F2265	F2330	F1560	F1718	-	-	-	-	-	-	-	-
<b>GPC21</b>	<b>F1876</b>	<b>F1675</b>	<b>F1892</b>	<b>F2006</b>	-	-	-	-	-	-	-	-
<b>GPC22</b>	<b>F1298</b>	<b>F1134</b>	<b>F1394</b>	<b>F1329</b>	<b>F1072</b>	<b>F853</b>	-	-	-	-	-	-
<b>GPC23</b>	<b>F2136</b>	<b>F2161</b>	<b>F2193</b>	<b>F2200</b>	-	-	-	-	-	-	-	-
GPC24	F1813	F1638	F2315	F2277	F2245	F1581	-	-	-	-	-	-
GPC25	F1994	F1883	F2110	F2031	F1937	F1808	-	-	-	-	-	-
Prioritised features of WWTP influent samples in negative ionisation mode												
<b>GPC1</b>	<b>F2846</b>	<b>F2858</b>	<b>F2893</b>	<b>F2900</b>	<b>F2962</b>	<b>F3009</b>	<b>F3015</b>	-	-	-	-	-
<b>GPC2</b>	<b>F2390</b>	<b>F2542</b>	<b>F2681</b>	<b>F2734</b>	<b>F3103</b>	<b>F3151</b>	<b>F3202</b>	-	-	-	-	-
<b>GPC3</b>	<b>F2745</b>	<b>F2841</b>	<b>F3003</b>	<b>F3039</b>	<b>F3040</b>	<b>F3048</b>	-	-	-	-	-	-
<b>GPC4</b>	<b>F2825</b>	<b>F2837</b>	<b>F2965</b>	<b>F2973</b>	<b>F3158</b>	<b>F3264</b>	-	-	-	-	-	-
<b>GPC5</b>	<b>F2400</b>	<b>F2616</b>	<b>F2864</b>	<b>F2883</b>	<b>F2958</b>	-	-	-	-	-	-	-
<b>GPC6</b>	<b>F2706</b>	<b>F2929</b>	<b>F3063</b>	<b>F3064</b>	-	-	-	-	-	-	-	-
<b>GPC7</b>	<b>F2448</b>	<b>F2531</b>	<b>F2576</b>	-	-	-	-	-	-	-	-	-
<b>GPC8</b>	<b>F2532</b>	<b>F2755</b>	-	-	-	-	-	-	-	-	-	-

## Chapter 4 Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches

Table-A 4-10 – List of *m/z*, retention time and the trend of prioritised features using GPCA of positive and negative ionisation mode (cf. Table S-9).

<i>Feature ID</i>	<i>Trend</i>	<i>m/z</i>	<i>Retention time in min</i>
<b>Prioritised features of positive ionisation mode</b>			
F1457	decreasing	281.1128	7.01
F1459	decreasing	281.1132	7.90
F1618	decreasing	313.1390	8.34
F1687	decreasing	335.1214	8.90
F1778	decreasing	363.1161	6.53
F1686	decreasing	335.1212	8.34
F1713	decreasing	341.1337	6.55
F1714	decreasing	341.1338	7.57
F1617	decreasing	313.1389	8.92
F1717	decreasing	343.1497	6.34
F1538	decreasing	299.1237	5.58
F924	decreasing	173.0209	4.54
F2062	increasing	476.3063	8.94
F1978	increasing	432.2800	8.36
F2131	increasing	520.3326	9.46
F1870	increasing	388.2535	7.72
F2223	increasing	564.3586	9.92
F1808	increasing	371.2275	7.71
F1937	increasing	415.2540	8.36
F1719	increasing	344.2275	6.97
F2031	increasing	459.2804	8.94
F2268	increasing	608.3848	10.34
F1612	decreasing	312.2532	17.91
F2121	decreasing	511.5189	19.06
F2145	decreasing	528.3640	11.61
F1935	decreasing	415.2538	6.91
F2243	decreasing	585.5196	19.06
F2166	decreasing	537.5346	19.21
F1616	decreasing	312.3260	20.74
F1747	decreasing	353.3158	16.99
F1809	decreasing	371.2275	6.21
F1581	decreasing	306.2400	17.91
F733	decreasing	153.0909	17.91
F913	decreasing	170.1175	17.89
F1232	decreasing	228.1953	19.72
F742	decreasing	154.1224	17.91
F916	decreasing	171.1015	18.15
F1429	decreasing	270.2786	19.54
F751	decreasing	155.1066	17.89
F728	decreasing	152.1067	17.90
F1286	decreasing	237.0866	10.02
F1214	decreasing	222.0760	5.58
F959	decreasing	178.0496	5.58
F960	decreasing	178.0498	7.98
F1381	decreasing	259.0689	10.02
F227	increasing	100.0755	4.88
F1529	increasing	298.8861	1.12
F2225	decreasing	567.5820	20.00
F2155	decreasing	533.5019	19.05
F1716	decreasing	341.2428	10.65
F2147	decreasing	529.4617	22.21
F2185	decreasing	550.3464	11.63
F2245	decreasing	589.5643	20.02
F2169	decreasing	539.5505	19.99
F935	decreasing	175.0152	8.88
F1240	decreasing	231.0781	8.88
F1349	decreasing	251.0468	8.18
F1097	decreasing	198.1277	5.42
F192	decreasing	98.9840	2.12
F1020	decreasing	186.0858	3.67
F398	decreasing	125.0153	3.67

Chapter 4 Evaluation of Non-Target Long-term LC-HRMS Time Series Data using Multivariate Statistical Approaches

Table-A 4-11 continued.

F1006	decreasing	184.0887	3.67
F92	decreasing	84.0557	2.20
F570	decreasing	137.0963	17.89
F1032	decreasing	186.1487	17.89
F1896	decreasing	399.3941	16.17
F1917	decreasing	408.3081	19.71
F1913	decreasing	407.3238	16.99
F1488	decreasing	286.3008	19.99
F1574	decreasing	304.2999	14.86
F1839	decreasing	380.2765	18.90
F2329	decreasing	680.4804	21.95
F1440	decreasing	276.1443	16.44
F2030	decreasing	459.2802	7.48
F2109	decreasing	503.3064	8.02
F1876	decreasing	389.3372	17.93
F1675	decreasing	330.2641	17.91
F1892	decreasing	398.2144	16.43
F2006	decreasing	445.4000	19.74
F1298	decreasing	240.1573	9.53
F1134	decreasing	204.9793	2.03
F1394	decreasing	263.0211	2.03
F1329	decreasing	247.2415	20.56
F853	decreasing	165.1122	17.89
F2136	increasing	523.0462	5.27
F2161	increasing	537.0352	5.26
F2193	increasing	552.0001	5.27
F2200	increasing	553.0063	5.26
<b>Prioritised features of negative ionisation mode</b>			
F2846	decreasing	216.0328	12.86
F2858	decreasing	218.0301	12.84
F2893	decreasing	232.0277	12.70
F2900	decreasing	234.0251	12.71
F2962	decreasing	249.9941	12.78
F3009	decreasing	269.9033	6.29
F3015	decreasing	271.9005	6.29
F2390	decreasing	77.0396	3.53
F2542	decreasing	121.0293	3.54
F2681	decreasing	165.0189	3.53
F2734	decreasing	179.0346	5.85
F3103	decreasing	321.2193	18.76
F3151	decreasing	353.0278	3.54
F3202	decreasing	383.9581	3.53
F2745	decreasing	187.0509	4.33
F2841	decreasing	214.0441	9.19
F3003	decreasing	269.0232	4.32
F3039	decreasing	293.1783	18.91
F3040	decreasing	294.0002	9.15
F3048	decreasing	295.9976	9.16
F2825	decreasing	209.0942	10.08
F2837	decreasing	212.9862	2.19
F2965	decreasing	250.9260	10.90
F2973	decreasing	252.9234	10.95
F3158	decreasing	353.2002	18.15
F3264	decreasing	529.4613	23.67
F2400	decreasing	78.9587	2.09
F2616	decreasing	153.0318	2.11
F2864	decreasing	221.1180	11.29
F2883	decreasing	229.0626	8.84
F2958	decreasing	249.0316	8.14
F2706	decreasing	171.0115	2.72
F2929	decreasing	243.0692	11.34
F3063	decreasing	299.1317	13.35
F3064	decreasing	299.1320	10.98
F2448	decreasing	93.0344	6.08
F2531	decreasing	118.0296	5.55

Table-A 4-11 continued.

F2576	decreasing	137.0242	6.10
F2532	decreasing	118.0297	4.38
F2755	decreasing	193.0997	12.25

#### 4.6.4 References

- Ahmad, I., Tang, D., Wang, T., Wang, M., Wagan, B., 2015. Precipitation trends over time using Mann-Kendall and spearman's Rho tests in swat river basin, Pakistan. *Adv. Meteorol.* 2015. <https://doi.org/10.1155/2015/431860>
- Camacho, J., 2011. Observation-based missing data methods for exploratory data analysis to unveil the connection between observations and variables in latent subspace models. *J. Chemom.* 25, 592–600. <https://doi.org/10.1002/cem.1405>
- Camacho, J., García-Giménez, J.M., Fuentes-García, N.M., Maciá-Fernández, G., 2019. Multivariate Big Data Analysis for intrusion detection: 5 steps from the haystack to the needle. *Comput. Secur.* 87, 101603. <https://doi.org/https://doi.org/10.1016/j.cose.2019.101603>
- Camacho, J., Garcia-Teodoro, P., Macia-Fernandez, G., 2017. Traffic monitoring and diagnosis with multivariate statistical network monitoring: a case study. *Proc. - 2017 IEEE Symp. Secur. Priv. Work. SPW 2017 2017-Decem*, 241–246. <https://doi.org/10.1109/SPW.2017.11>
- Camacho, J., Rodríguez-Gómez, R.A., Saccenti, E., 2017. Group-Wise Principal Component Analysis for Exploratory Data Analysis. *J. Comput. Graph. Stat.* 26, 501–512. <https://doi.org/10.1080/10618600.2016.1265527>
- Camacho, J., Theron, R., Garcia-Gimenez, J.M., Macia-Fernandez, G., Garcia-Teodoro, P., 2019. Group-Wise Principal Component Analysis for Exploratory Intrusion Detection. *IEEE Access* 7, 113081–113093. <https://doi.org/10.1109/access.2019.2935154>
- Jolliffe, I.T., 2002. *Principal Component Analysis*. Springer-Verlag New York, New York. <https://doi.org/https://doi.org/10.1007/b98835>
- Jolliffe, I.T., Trendafilov, N.T., Uddin, M., 2003. A Modified Principal Component Technique Based on the LASSO. *J. Comput. Graph. Stat.* 12, 531–547. <https://doi.org/10.1198/1061860032148>
- Lamchin, M., Lee, W.K., Jeon, S.W., Wang, S.W., Lim, C.H., Song, C., Sung, M., 2019. Corrigendum to “Mann-Kendall Monotonic Trend Test and Correlation Analysis using Spatio-temporal Dataset: the case of Asia using vegetation greenness and climate factors” (*MethodsX* (2018) 5 (803–807), (S2215016118301134), (10.1016/j.mex.2018.07.006)). *MethodsX* 6, 1379–1383. <https://doi.org/10.1016/j.mex.2019.05.030>
- Mondal, A., Kundu, S., Mukhopadhyay, A., 2012. 70 Rainfall trend analysis by Mann-Kendall test: A case study of north-eastern part of cuttack district, orissa. *Online) An Online Int. J. Available* 2, 70–78.
- Shadmani, M., Marofi, S., Roknian, M., 2012. Trend Analysis in Reference Evapotranspiration Using Mann-Kendall and Spearman's Rho Tests in Arid Regions of Iran. *Water Resour. Manag.* 26, 211–224. <https://doi.org/10.1007/s11269-011-9913-z>

Sharma, D., Kumar, B., Chand, S., 2019. A Trend Analysis of Machine Learning Research with Topic Models and Mann-Kendall Test. *Int. J. Intell. Syst. Appl.* 11, 70–82. <https://doi.org/10.5815/ijisa.2019.02.08>

Yue, S., Pilon, P., Cavadias, G., 2002. Power of the Mann–Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series. *J. Hydrol.* 259, 254–271. [https://doi.org/https://doi.org/10.1016/S0022-1694\(01\)00594-7](https://doi.org/https://doi.org/10.1016/S0022-1694(01)00594-7)

## Chapter 5 Spatial Trend Detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

### 5.1 Introduction

Wastewater treatment is a process that converts contaminated wastewater into wastewater that can be recycled or directly reused with minimal environmental impact. Processes in industrial wastewater treatment plants (WWTP) involve mostly three treatment steps employing a combination of mechanical, biological and chemical treatment methods (Özkaraova et al., 2018). In general, the subsistence of industrial WWTPs is cost-intensive in terms of plant space, equipment and labour (Harmon et al., 2011; Ranade and Bhandari, 2014). Therefore, assessing the performance of the WWTP is essential, regarding efficiency in treatment processes. Today, the use of liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) is the most powerful technique for screening approaches of trace organic compounds (TrOCs) and their transformation products at environmentally relevant concentrations in complex environmental samples (Haun et al., 2013; Hermes et al., 2018; Kiefer et al., 2019; Leendert et al., 2015; Park and Snyder, 2020; Schollée et al., 2016; Yan et al., 2017). A powerful technique for detecting and identifying emerging TrOCs in the environment is the non-target screening (NTS). As manufacturing activities of industrial processes are mostly undisclosed because of production data confidentiality, because further chemical reactions could occur during transport in the wastewater transport system and subsequent treatment, substance information is rarely available. Therefore, compounds present in production wastewater could remain unknown. Target screening approaches are likely to miss specific compounds that arise during the production, including intermediates, products, catalysts and synthesis by-products. However, NTS based on LC-HRMS full scan measurements enables both the detection of known and unknown TrOCs. Therefore, NTS helps to expand the assessment of industrial wastewater regarding the efficiency in treatment processes (Bader et al., 2017).

LC-HRMS offers the opportunity to comprehensively assess water treatment by comparing the signal heights of all detectable compounds during treatment, without the preselection of previously known compounds. For example, Bader et al. (2017) introduced a suitable assessment of drinking water treatment by NTS using LC-HRMS. In that study, the analytical window was opened for treatment evaluation by detecting all compounds which are separable by the applied chromatography and ionisable by the used ionisation method. Even for signals whose identity remained unknown during an evaluation, the features (mass-to-charge ratio,

retention time and intensity) could be tracked during the treatment process to assess the process. Further studies, already applied NTS by LC-HRMS to evaluate the elimination efficiency of WWTPs (Itzel et al., 2020; Nürenberg et al., 2015; Parry and Young, 2016). Therefore, as demonstrated, NTS is already implemented in assessing WWTPs, providing a more holistic picture of the entire process and less bias caused by preselection of known substances. An identification is obligatory for the evaluation of the treatment process. Mostly, correlations can be recognised and triggers can be tracked by analysing just the features of the monitored sampling sites without structure knowledge (Bader et al., 2017). However, in most cases, the identification of relevant features is in demand (Itzel et al., 2020). A significant issue of NTS is the management and processing of the massive amount of information and data that is produced per sample, especially when subsequent identification is desired and in the case of (industrial) wastewater samples (see chapter 4). To solve this high-throughput of data analysis, prioritisation methods are applied to process all chromatographic and accurate mass information. Therefore, the data set is usually transformed into a lower-dimension from which conclusions could be extracted (Hedgspeth et al., 2019; Hohrenk et al., 2019; Krauss et al., 2019). Trend analyses enable both, prioritisation of relevant features for identification and the assessment of WWTP processes, both of which were already applied in recent studies (Alygizakis et al., 2019; Ruppe et al., 2018; Schlüsener et al., 2015). Spatial trend analyses were already implemented by Bader et al. (2017) comparing process influent and effluent samples and by Krauss et al. (2019) applying LC-HRMS with NTS for the prioritisation of site-specific contamination, aiming at the identification of the underlying chemicals (Krauss et al., 2019). However, these approaches were focused on selective samples at single time points to assess the treatment process, while the assessment of the process over time was not investigated. Previously, other studies combined spatial and temporal trend analysis assessing water quality but did not use NTS to determine spatial and temporal water quality variations (Barakat et al., 2016; Singh et al., 2004; Zeinalzadeh and Rezaei, 2017). Consequently, this study aimed to combine the NTS long-term trend analysis, introduced in chapter 4, with spatial trend analysis across five sampling sites of an industrial WWTP. Therefore, for the assessment of the WWTP, the spatial based evaluation scheme of Bader et al. (2017) is used in combination with the spatial analysis of Krauss et al. (2019), which expresses site-specific contamination in a single value by rarity scores. Both studies are further expanded by the trend analysis of chapter 4 to enable the assessment over time. The results are expected to help evaluating the temporal behaviour of an industrial WWTP. In the presented study data treatment strategies were applied (1) to determine the similarities and differences between the sampling sites to build site-specific fingerprints for expanding monitoring, (2) to evaluate the temporal contribution variations in wastewater treatment quality and (3) to prioritise features for later obligatory identification of substances that affect the treatment quality.



## 5.2 Experimental Section

### 5.2.1 Chemicals and Reagents

All solvents used in the present work were of LC-MS grade. Methanol was from Honeywell Riedel-de-Haën™ (Seelze, Germany), a Milli-Q® (Q-PoD®) ultra-pure water system from Merck KGaA (Darmstadt, Germany) was used and formic acid 99% was purchased from Fisher Chemical (Geel, Belgium). All internal standards (ISTD) were of a quality grade suitable for trace analysis, purity  $\geq 95\%$ . The corresponding chemical abstract service (CAS) numbers of used analytes are listed in the Appendix (see 5.6).

### 5.2.2 Sampling Sites and Storage

Details of monitored water matrices are shown in Table 5-1 and a schematic overview of studied sampling sites is shown in chapter 3 of this dissertation. 24-h composite flow industrial wastewater samples were used based on three relevant matrices: raw influent wastewaters (SP 1, SP 2, SP 3), wastewaters sampled within the WWTP during the biological treatment (SP 4) and WWTP effluents (SP 5). The WWTP is designed with primary sedimentation, activated sludge processes with biological nitrogen and phosphorus removal and secondary sedimentation. Acid raw influent wastewater is neutralised before it enters treatment processes to protect the microbial organisms. The sampled WWTP treated both industrial and municipal wastewater. The samples were collected from November 2018 to March 2019 within the routine sampling. The samples were stored in pre-cleaned high-density polyethylene (HDPE) bottles and transported to the laboratory for analysis. If the direct analysis was not possible, aliquots (1 mL) were stored in the freezer ( $-22\text{ }^{\circ}\text{C}$ ) until analysis.

Table 5-1 – Table of different sampling sites (SP) across the treatment processes of the WWTP during November 2018 to March 2019 used for spatial-temporal trend analysis. The corresponding schematic overview is shown in chapter 3 of the dissertation.

<b>ID</b>	<b>Matrix</b>	<b>Number of Samples</b>
<b>SP 1</b>	Municipal wastewater	54
<b>SP 2</b>	Composite wastewater of industrial area	65
<b>SP 3</b>	Composite industrial wastewater directly before the WWTP	69
<b>SP 4</b>	Wastewater within the WWTP during the biological treatment	74
<b>SP 5</b>	The WWTP effluent	65

### 5.2.3 Trend Detection Workflow

The untargeted trend detection workflow to assess the industrial WWTP was implemented in four steps: (I) LC-HRMS acquisition, (II) data processing, (III) assessment of wastewater treatment processes and (IV) an identification of TrOCs (cf. Figure 5-1).

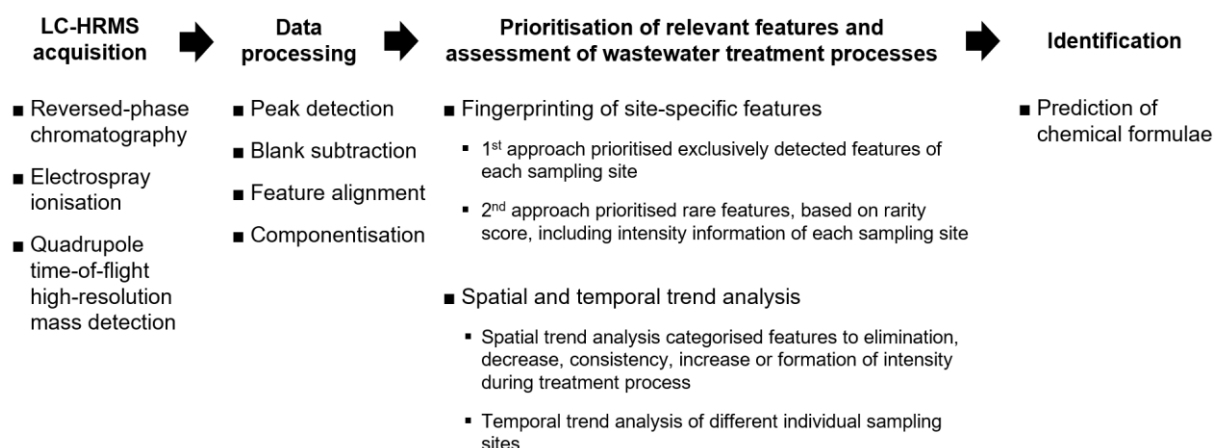


Figure 5-1 – Flowchart of the trend detection workflow applied for the assessment of the industrial WWTP based on spatial and temporal treatment behaviour.

#### 5.2.3.1 LC-HRMS Acquisition

Detailed LC-HRMS acquisition method is described in section 5.6.1.2 of the Appendix. Samples were diluted by a factor of 10 with mobile phase A and spiked with an internal standard solution containing 13 isotope-labelled standards (ISTD, see Appendix 5.6). Wastewater samples of SP 1 to SP 3 were centrifuged for one minute at 2500 rpm (Centrifuge, 400 Function Line, Heraeus, Hanau, Germany) before dilution. 5 µL of the diluted and spiked wastewater samples were injected in the LC, respectively. The chromatographic separation was performed by using an Agilent 1290 Infinity HPLC system (Agilent Technologies, Waldbronn, Germany) with Raptor™ Ultra Aqueous C18 analytical column (100 mm x 2.1 mm, 3.0 µm particle size, Restek GmbH, Bad Homburg v. d. Höhe, Germany) with Restek™ Trident

cartridge (10 x 2.1 mm) and a filter (2 mm, 0.5  $\mu\text{m}$ ; Restek GmbH, Bad Homburg v. d. Höhe, Germany). For detection, samples were measured (in sequence) in positive and negative electrospray ionisation mode (Turbo V<sup>TM</sup> of SCIEX GmbH, Darmstadt, Germany) on the hybrid quadrupole time-of-flight mass spectrometer x500R qTOF (SCIEX GmbH, Darmstadt, Germany). In the TOF-MS survey of 100 ms up to 10 data-dependent TOF-MS/MS scans in each cycle (0.56 s, intensity threshold of 50 cps) covering a  $m/z$  70 – 950 were produced. Ions, as well as their isotopes, were excluded from data-dependent acquisition for a period of 4 s after three occurrences.

#### 5.2.3.2 Data Processing

MarkerView (MV, Version 1.3.1, SCIEX, Framingham, USA) was used for peak detection, peak alignment across different samples, normalisation of retention times and blank value subtraction. The specific parameters were optimised, keeping the number of false negatives to a minimum (see Appendix 5.6.1.3). Blank peaks were subtracted using overall exclusion to remove background ions and chemical noise introduced by the lab procedures. The average values of peak intensities and retention times of ISTD (see 5.6.1.1) were used to correct the retention times. Furthermore, points without data information in the data matrix were replaced with signal noise of the instrument. MV converted data to centroid data in the form of peak retention times and peak  $m/z$  ratios. A data matrix of each sampling site was built with listed intensities of detected features against monitored samples (time points). Subsequently, the generated data matrices of MV were imported into the MATLAB<sup>®</sup> environment (release R2017b, MathWorks, Inc., Natick, USA). Further processing steps were performed using an in-house data treatment script. The initial matrices were converted into a combined matrix of data which is suitable for multivariate data analysis. The aligned features over all sampling sites and all time points were presented in one data matrix by averaging the time over all sampled stations. Figure 5-2 schematically illustrates the built data matrix.

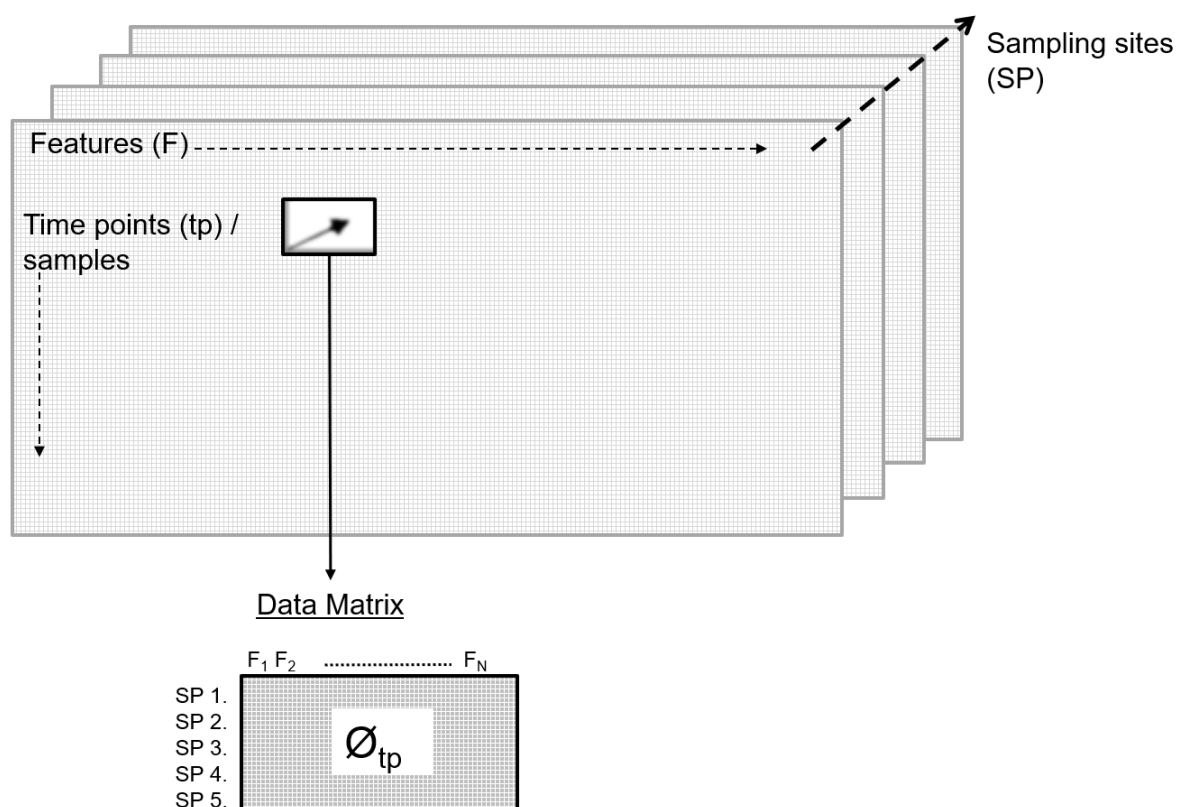


Figure 5-2 – Schematic representation of the multiple data collection for the used data matrix of all aligned features (F) of all sampling sites (SP) and measured time points (tp).

### 5.2.3.3 Assessment of Wastewater Treatment Processes

For assessing the treatment processes of the WWTP, spatial and temporal trend analysis was performed (cf. Figure 5-1). For the analysis of sampling sites, site-specific features were assigned by selecting those features exclusively detected at the individual sampling site. From the aligned data matrices of each sampling site,  $m/z$  ratios, retention times and all intensities over all measured samples of five months were used. Tolerances for exclusively determining site-specific features were a maximum mass error of  $\pm 2$  ppm and a maximum retention time error of  $\pm 0.05$  min. Furthermore, a maximum number of ‘non-detects’ or zeros of 10 per feature was acceptable, or else the features were excluded for site-specific features. Furthermore, principal component analysis (PCA) was performed on the sampling sites using the data matrix of Figure 5-2. Among several data pre-processing approaches, logarithmic transformation was selected. Thus, skewness was reduced, giving normally distributed variables. In addition, normalisation followed by autoscaling was applied due to different magnitudes and scales of features and also because of measuring wastewater samples over time.

Besides, site-specific features were assigned in a second approach using the rarity score introduced by Krauss et al. (2019). The rarity score prioritised those features which were

site-specific in case of significant differences in intensity among the samples (as a proxy for concentration) or by a reduced frequency of occurrence (Krauss et al., 2019). Equation I was used to determine the rarity score based on Krauss et al. (2019).

$$\text{Rarity score} = \frac{\text{maximum intensity across all sites}}{\text{median intensity across all sites}} \times \frac{\text{total number of sampling sites (5)}}{\text{number of positive detects}} \quad (I)$$

With ‘number of positive detects’ indicating positive detects of individual features at each sampling site. Furthermore, features were prioritised for assessment of the WWTP, which could be classified according to the categories introduced by Bader et al. (2017) during spatial based time trend analysis (see Table 5-2). The classification based on the fold change ( $f_c = \text{intensity}_{\text{Effluent}} / \text{intensity}_{\text{Influent}}$ ) between the effluent and influent samples across the treatment process. Thus, the WWTP could be assessed by comparing the intensities of the WWTP effluent and influent. The  $f_c$ s were classified to five distinct categories, as shown in Table 5-2. Each feature was classified as average intensity over time at the individual sampling site.

*Table 5-2 – Categories applied for the assessment of wastewater treatment processes. The calculated fold chain ( $f_c$ ) is based on feature’s intensity ( $f_c = \text{intensity}_{\text{Effluent}} / \text{intensity}_{\text{Influent}}$ ) across treatment process. Categories based on Bader et al. (2017).*

<b>Category</b>	<b>Fold chain (<math>f_c</math>) interval</b>
Elimination	$0.00 \leq f_c < 0.20$
Decrease	$0.20 \leq f_c < 0.50$
Consistency	$0.50 \leq f_c \leq 2.00$
Increase	$2.00 < f_c \leq 5.00$
Formation	$5.00 < f_c \leq \infty$

For temporal trend analysis, NTS long-term time series of sampling sites across the treatment process were evaluated. Feature prioritisation was based on the introduced time trend analysis workflow of chapter 4. In brief, features, which increased/decreased in intensity over time, were prioritised by multivariate analysis using feature-wise PCA and group-wise PCA (GPCA, temporal-wise) of the matrix. Initially, corresponding data pre-treatments (log-transformation, normalisation and autoscaling) were performed because of the skewness of the data from the normal distribution (checked by normality tests: the Shapiro-Wilk and the Kolmogorov-Smirnov test). Feature exploration and prioritisation were evaluated using row-wise PCA and calculation of  $Tscore_t$  values, which are a combination of the D-statistic and the Q-statistic values in a

single selecting score. Using GPCA, which is a recent sparse variant of PCA, every component contains non-zero loadings for a single group of correlated features. Thus, several groups of features can be detected and visualised, as shown recently (Camacho et al., 2019). Furthermore, a pre-filtering of the data matrix was executed by univariate statistics (Spearman's rank correlation) to reduce monitored data amount.

#### 5.2.3.4 Identification Procedures

In case of identification, proposed formulae were determined by the *SCIEX OS* software in 'Non-targeted Screening' workflow (Version 1.4.1, SCIEX GmbH, Darmstadt; Germany). Elements considered in the molecular assignment (and their minimum and maximum counts) included: C<sub>0-49</sub>, H<sub>0-75</sub>, N<sub>0-10</sub>, O<sub>0-16</sub>, P<sub>0-1</sub>, S<sub>0-3</sub>, F<sub>0-7</sub> and Cl<sub>0-5</sub> with a mass tolerance of 10 ppm. Online databases, e.g. *ChemSpider* (Royal Society of Chemistry, 2017) and the *High-Resolution Accurate Mass Libraries* (HRAM All-in-one v.1.1, contains spectra for 2231 compounds, SCIEX GmbH, Darmstadt; Germany) were used to identify compounds tentatively.

### 5.2.4 Quality Control

Instrumental drift was monitored by analysing a quality control sample (QC) repeatedly throughout the sample sequences. QC was produced consisting of 70 target analytes (cp. Appendix 5.6.1.1) with 100 µg L<sup>-1</sup> concentration and ISTD mixture of 13 isotope-labelled (deuterated) compounds. The QC standard was run for both ionisation modes at the beginning and the end of each sample sequence. Besides, within each sequence, a blank, consisting of 1 mL Milli-Q® with 0.1% formic acid, was measured to monitor background contaminations. Furthermore, the category of consistency (see section 1.2.2, III) served as quality control for the data analysis. The data analysis was assessed by monitor the spiked ISTD. In case of proper data analysis, the ISTDs were categorised in the category of consistency (see 5.2.3.2), as they were spiked with constant concentration to wastewater samples. Results are presented in Appendix 5.6.1.4. ISTD were all categorised in the category of consistency, except of few neglectable outliers.

## 5.3 Results and Discussion

### 5.3.1 Characterisation of Features of studied Sampling Sites

#### 5.3.1.1 Comparison of Sampling Sites

Features listed in Table 5-3 were extracted for individual sampling sites from November 2018 to March 2019, in the daily composite samples. These features represent the aligned data output of MV for all samples over time at the individual sampling sites, respectively. However, it may still include multiple peaks resulting from adducts and false positive integrations of signal noise (Hug et al., 2014). Therefore, the implemented componentisation strategy of chapter 4 using GPCA was applied. As a result, the number of detected features was substantially reduced, as shown in Table 5-3 (cf. Figure 5-3). Each of these features at every sampling site and time point was assigned a representative signal intensity that can be used as a proxy of concentration but excluding matrix effects in electrospray ionisation (in particular ion suppression) that may affect this relation (Schollée et al., 2016). To reduce the impacts of matrix effects on signal intensities for the different matrices, wastewater samples were diluted by a factor of 10 with pure-water mixed with 0.1% formic acid (see also chapter 3). The suitability of this approach was demonstrated previously in the literature (Nürenberg et al., 2015).

Table 5-3 – Aligned features initially detected within five months and after applying componentisation by GPCA introduced in chapter 4.

<b>Sampling sites</b>	<b>Aligned features of five months</b>	<b>Aligned features after componentisation by GPCA</b>
SP 1	666	435
SP 2	4926	2596
SP 3	3303	1898
SP 4	4178	2959
SP 5	2387	1043

The number of features in wastewater samples was reduced across the treatment process, as shown in Figure 5-3. However, there was a small increase of features at the sampling site within the WWTP (SP 4) compared to the WWTP influent samples (SP 3). The increase could be explained, in agreement with the literature, by transformation reactions during biological treatment steps (Deeb et al., 2017; Huntscha et al., 2014; Itzel et al., 2020; Verlicchi et al., 2012). Besides, the bulk of features appeared at this sampling site in the lower retention time and  $m/z$  region, which confirmed the previous suggestion, as transformation products are

commonly more mobile than the parent compounds, as shown for several pesticides in the literature (Buttiglieri et al., 2009; Richardson and Kimura, 2016).

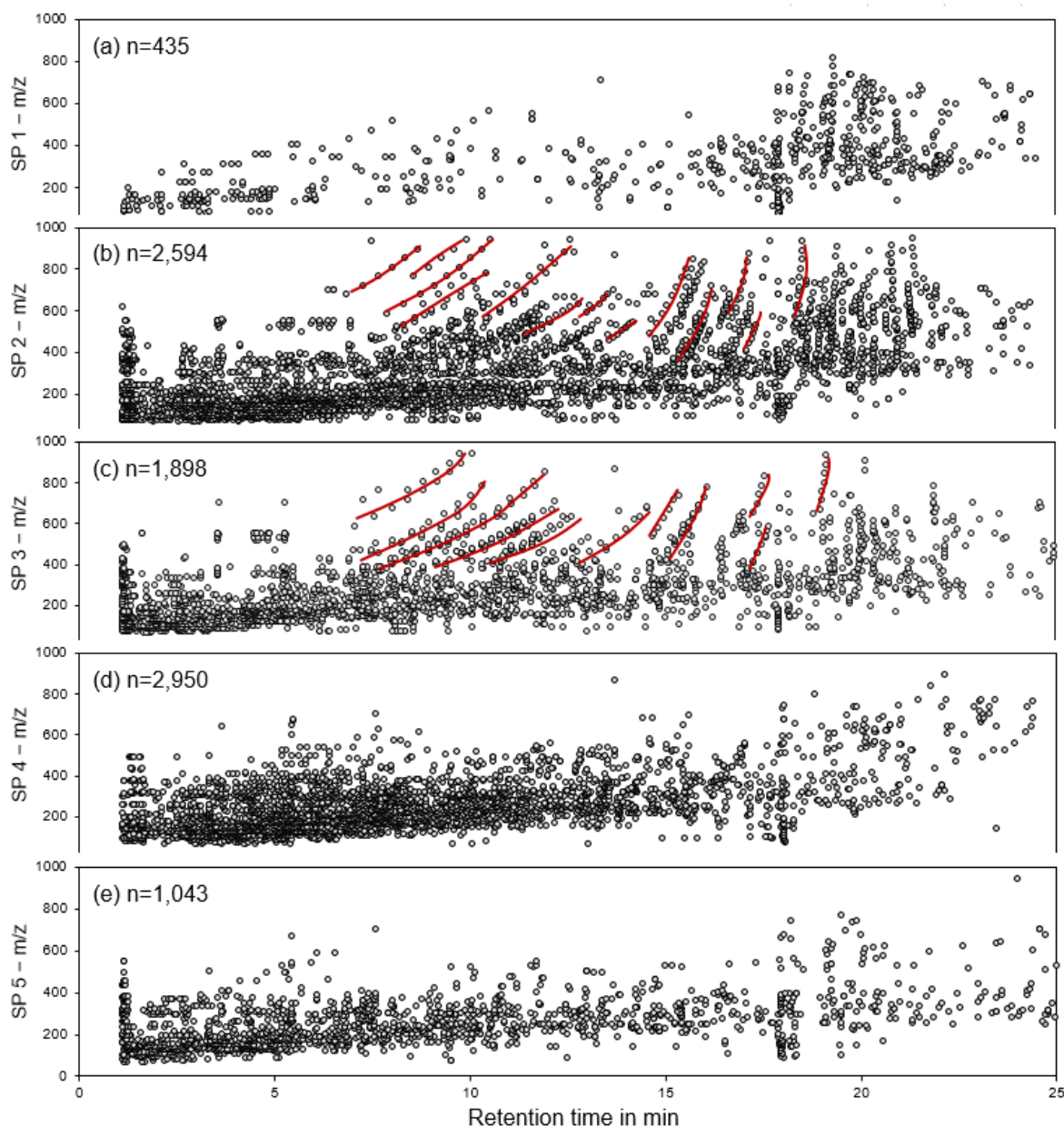


Figure 5-3 – Mass-to-charge ratio ( $m/z$ ) vs retention time scatter plots of aligned features by NTS over five months. In (a) the municipal wastewater (SP 1), in (b) the composite wastewater of the industrial area (SP 2), in (c) the composite industrial wastewater directly before the WWTP (SP 3), in (d) the wastewater within the WWTP (SP 4) and in (e) the WWTP effluent (SP 5) are shown. In (b) and (c) some homologous series are highlighted by red lines.

Homologous series, which differ in a typical chemical subunit resulting in similar changes of  $m/z$  ratio and retention time, occurred at sampling sites SP 2 and SP 3 (cf. Figure 5-3 b and c). They occur in natural organic matter and classes of anthropogenic substances like surfactants, polyfluorinated compounds or chlorine substitution series (Jobst et al., 2013). They are of analytical interest and part of several research studies in NTS (Loos and Singer, 2017; Schollée et al., 2015; Verkh et al., 2018). However, compounds of the homologous series were



eliminated in the studied industrial WWTP as they did not occur in sampling sites SP 4 and SP 5 anymore (see Figure 5-3 d and e). The number of features detected in municipal wastewater (SP 1) was comparatively low, reflecting the sample load and consequently the importance of the monitoring of industrial wastewater.

Figure 5-4 shows the resulting score plot of PCA analysing sampling sites. The data matrix described in Figure 5-2 served as the basis for the analysis, in which the intensities were listed as an average over the time points of all the features against the sampling sites. The first two principal components (PC1 34.0% and PC2 31.1%) provided a general view on the variations of studied sampling sites describing 65.1% of the data variation. WWTP influent samples of SP 2 and SP 3 were clustered, as well as SP 1 and SP 5 (see Figure 5-4). In contrast, the sampling site within the WWTP of biological treatment was different from others, confirming the previous observations.

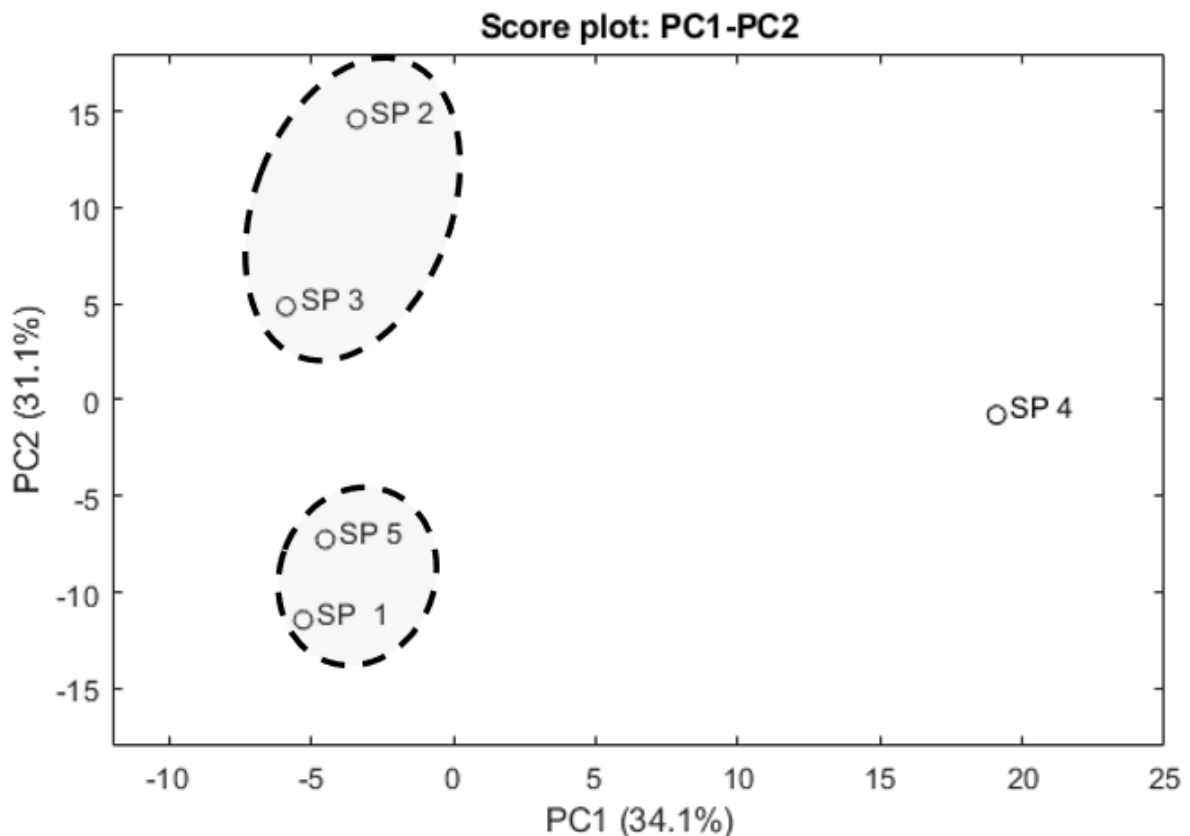


Figure 5-4 – Score plot of the PCA of the data matrix having sampling sites in the row, samples in the column and the intensity of features averaged over five months.

Sampling sites SP 2 and SP 3 were described as a closed system (see chapter 3, Figure 3-1), no external effects happened in between, as they were combined by one pipe of wastewater. Thus, samples taken from these sampling sites could only differ because of chemical reactions during transport within the wastewater pipe network. However, the results demonstrated that

such processes had no significant effect on feature patterns. The clustering of SP 1 and SP 5 indicated structural similarities in feature pattern between municipal wastewater and the WWTP effluent. Therefore, the relative load of industrial wastewater in the WWTP was comparatively low. The similarity of the feature pattern indicates that the dominant number of features came from municipal wastewater. Thus, by applying NTS, more features were detected which were comparatively poorly degradable in the municipal wastewater, while for industrial wastewater, this was less the case. Consequently, more features of the municipal wastewater could be detected in the WWTP outlet.

#### 5.3.1.2 Fingerprinting of Site-Specific Features

At each sampling site, site-specific unknown compounds were found by selecting in a first approach features detected at only one sampling site. This systematic comparison of wastewater samples resulted in 15 features exclusively detected in SP 1, 33 features in industrial wastewater (SP 2), 21 features in SP 3, 47 features within the WWTP (SP 4) and 9 features in the WWTP effluent (SP 5). The examination of the prioritised features over time resulted in both, features with no trend, which could represent background features corresponding to the sampling sites and features having fluctuation trends of intensity over time, which could correspond to production peaks of the industrial area. Two features of SP 4 had an increasing trend in intensity over time ( $m/z$  151.0309, 13.02 min and  $m/z$  262.111, 9.76 min). Details of prioritised features are listed in the Appendix (see 5.6.2.1, Table-A 5-4).

In a second approach, features were prioritised by using the rarity score of equation I. This approach considered the intensity related occurrence of features. The score was calculated by the maximum intensity of one feature compared to its median intensity in all samples, taking the number of positive detects in the individual sampling site into account (e.g. for feature with  $m/z$  234.0723 and 7.12 min the rarity score was  $\frac{3.90E+06 \text{ cps}}{7.52E+04 \text{ cps}} \times \frac{5}{1} = 259.22$ ). By defining a threshold for the rarity score, features were scaled based on their intensity related occurrence, which can be used for the determination of site-specificity. Krauss et al. (2019) reported that the magnitude of rarity scores depended on the used instrument and the data set itself. Therefore, the threshold for prioritisation was determined for this study individually (see Figure 5-5).

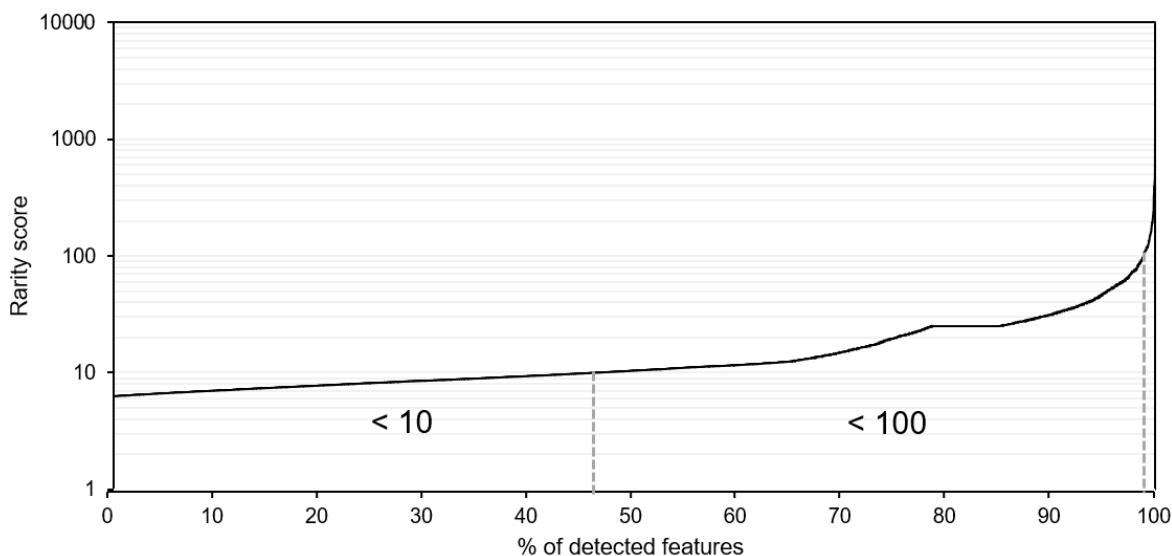


Figure 5-5 – Rarity score of all detected features (8,931 in total, see Table 5-1) based on Krauss et al. (2019). The scores are sorted increasingly on a logarithmic scale.

47% of the resulting rarity scores were between 0 and 10, 52% of the determined rarity scores were between 10 and 100 and only 1% of the detected features had scores above 100. Features having rarity scores below 10 had small intensities ( $< 6.00E+04$  cps) or were frequently present at sampling sites. Therefore, they were insignificant assessing the studied WWTP and prioritising site-specific features. High rarity scores, in contrast, indicated features having higher intensities or were rarely present at studied sampling sites. Thus, features having a rarity score above 100 (1%) were prioritised (see Figure 5-5). At this level, 35 features were selected down to a signal intensity of  $1.00E+06$  cps with a median number of positive detects of two sampling sites. Thus, they were classified as rare peaks for monitored sampling sites. The prioritised features were tentatively identified as level 4 candidates (Schymanski et al., 2014). Features with corresponding  $m/z$  ratios, retention times and predicted sum formulae are listed in Table-A 5-5.

In general, the determination of site-specific features with both approaches could be used as source-related contamination fingerprints for further investigation (Krauss et al., 2019). By monitoring their intensity over time, the specific production pattern could be assigned. For example, in online monitoring, the rarity score could be continuously calculated and plotted, as shown in Figure 5-5, to assess the treatment efficiency of the WWTP. LC-HRMS using NTS could be implemented, automatically determining the rarity score of all detected features. It would be conceivable that operators of WWTPs set a threshold for a rarity score of detected features, which could be used to control the WWTP. Therefore, the rarity score might offer a possibility to monitor a WWTP in online monitoring. Furthermore, the fingerprint of a sampling site defined as a combination of NTS features may help to understand complex contamination

patterns in industrial wastewater not only as mixtures of individual compounds but as an overlay of source-related features. This can be used to estimate, prioritise and adapt contributions of pollution sources (Brack et al., 2018). Furthermore, the fingerprint of sampling sites will vary in case of incidents in the WWTP. Thus, for instance, events during the treatment processes could be assigned to features of one sampling site, which in series corresponded to individual production plants in the industrial area.

### **5.3.2 Assessment of the WWTP using Trend Analysis**

For the assessment of the WWTP using trend analysis, two approaches were followed based on spatial and temporal analysis expanding to unknown TrOCs. The first approach focused on spatial trend analysis of average time results which extracted features showing elimination, decrease, consistency, increase or formation in their intensity during the treatment processes. The second approach was focused on temporal trends over different sampling sites to assess the WWTP by determining features showing relevant time pattern.

#### *5.3.2.1 Spatial Trend Analysis*

The spatial trend analysis was based on Bader et al. (2017). The concept introduced for assessing processes of drinking water treatment was applied to assess wastewater treatment processes and was extended by time-series measurements. Features detected over time at the individual sampling sites were categorised between process effluent and influent, as demonstrated by Bader et al. (2017). Figure 5-6 shows exemplarily the features between sampling sites SP 5 and SP 3. The categorisation was based on the calculated fold change of intensity to elimination, decrease, consistency, increase and formation during the treatment process. The detailed analysis of the process is shown in Figure-A 5-2 (see Appendix 5.3).

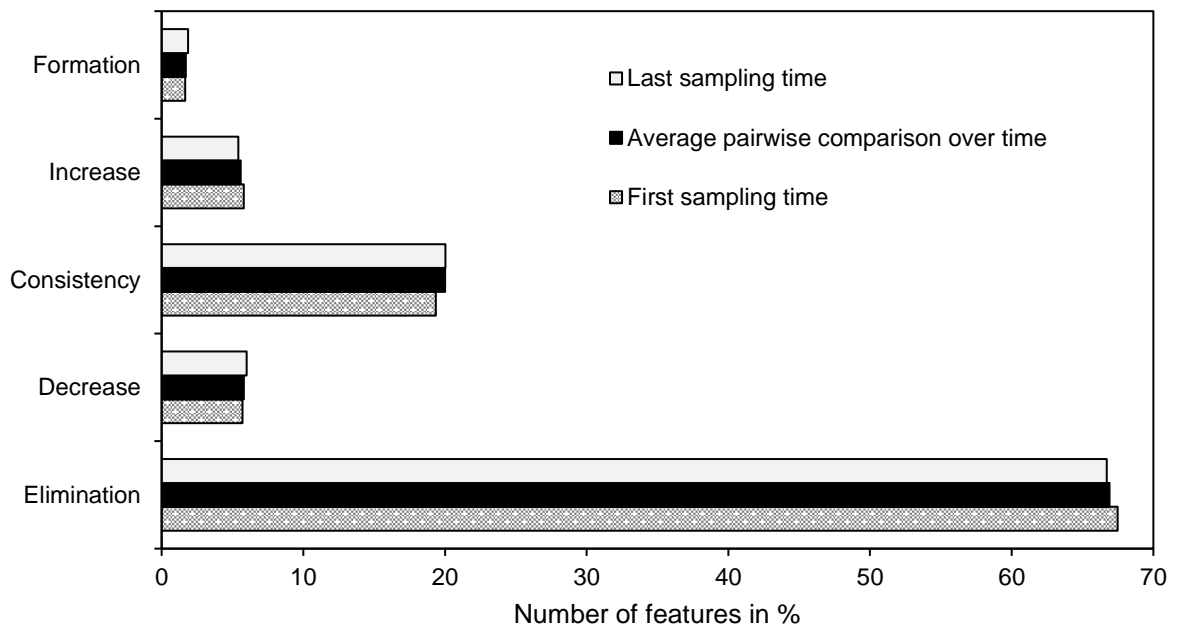


Figure 5-6 – Result of treatment efficiency of the WWTP. The comparison of the process WWTP effluent and influent of samples taken over five months is shown for the first sampling time, the average and for the last sampling time. Evaluation is based on Bader et al. (2017). The quotient of SP 5/SP 3 was calculated by dividing the intensities of, e.g. first time point by the intensity of, e.g. second time point. Details of every sampling time are shown in the Appendix (see 5.6.1.3).

The result showed that the treatment process was not influenced by time (regarding samples of November 2018 to March 2019). The results of the other sampling sites are shown in Figure 5-7.

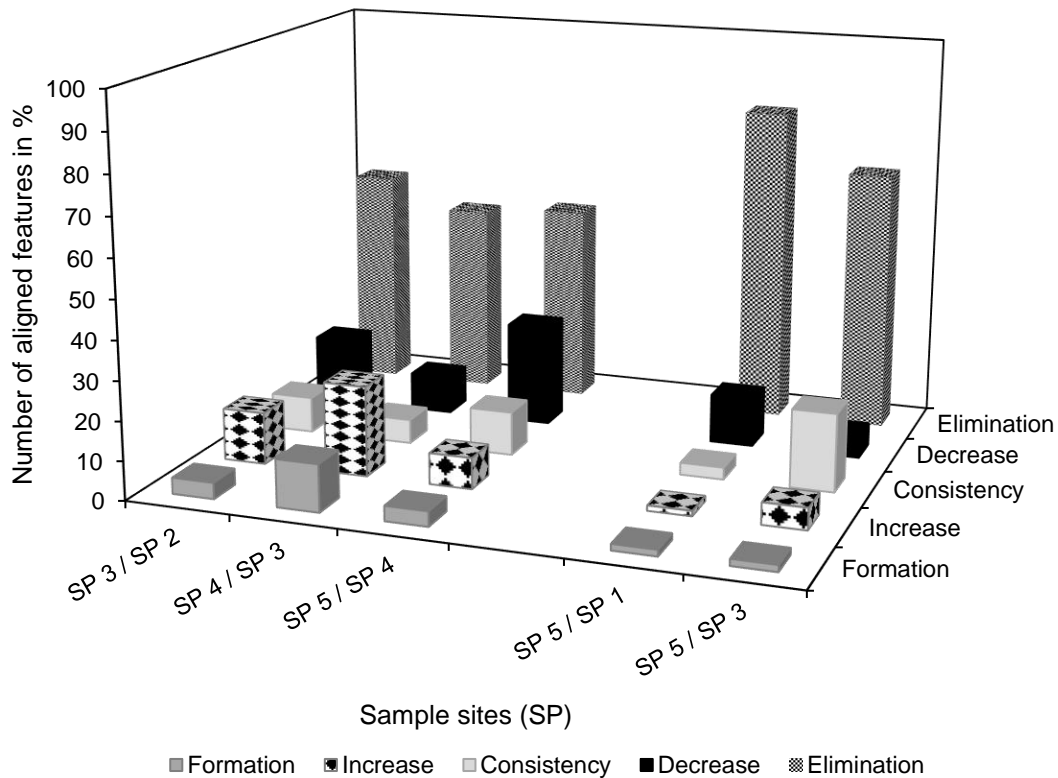


Figure 5-7 – Result of treatment efficiency of the WWTP of studied five months. The number of aligned features in the different categories was based on fold change determined by pairwise comparisons of sampling sites (SP), respectively.

For all sampling site comparisons, the category of elimination showed the highest number of features, followed by the categories of decrease, increase, consistency and at least formation. Most decrease and elimination of the intensity of features during treatment was observed between sampling sites SP 4 and SP 5, during the biological treatment step (cf. Figure 5-7). In comparison, the total decrease and elimination over the whole treatment process (from SP 3 to SP 5) was less distinct compared to biological treatment because of the small increase in the intensity of features between sampling sites SP 3 to SP 4. Therefore, the determination of more sampling sites than just the comparison of the overall process influent and effluent optimised the assessment of treatment efficiency of the studied WWTP. The influences of features' intensity could be assessed in more detail by monitoring sampling sites nearby every treatment step in the WWTP. To verify the category of formation,  $f_c$  was also calculated between the intensities of process influent sample of the studied day and the intensities of process effluent of the next day to check whether some features had a delay in flow time in the WWTP (cf. Figure-A 5-3). The results did not change in a significant way which may constitute there was no delay in flow time of features during the treatment process. Therefore, the quality of treatment within the WWTP was reached as expected. The decreasing and eliminating

intensity of features during treatment were the most intensive categories and the category of formation was in comparison less frequently found during the treatment process. Thus, the results demonstrate a proper treatment process of the monitored WWTP for unknown TrOCs. However, in contrast to Figure 5-4, which clustered SP 1 and SP 5 the category of consistency is rather low compared to the pairwise comparison of SP 3 and SP 5. The reason for this could be the applied thresholds for categories. For example, the category of elimination comprises fold changes until 0.2, which means there were still small intensities of features detectable in the WWTP effluent. In PCA the intensity had no significant influence because of the average inclusion. Therefore, in PCA the patterns of the sampling sites were analysed, while in Figure 5-7 the individual treatment processes and intensities of the features were considered.

### 5.3.2.2 *Temporal Trend Analysis*

A total of 1,898 features were initially detected in the industrial wastewater influent of the WWTP (SP 3, after componentisation using GPCA). 90.5% of these features were fully eliminated within the WWTP and not detected anymore in the WWTP effluent. Only 9.5% of the influent features remained, mostly at substantially reduced intensities (< 10% of the WWTP influent intensities, see the feature list in Appendix 5.6.2.3, Table-A 5-6). By applying time trend analysis of chapter 4, 28 features of those detected in both the WWTP influent and the WWTP effluent had a decreasing intensity trend over time, further 152 features of the WWTP effluent samples had other trends (fluctuation trends, peaks). However, fluctuations described normality because of different production cycles and fluctuating industrial discharges which caused short-term high peak concentrations in WWTPs (Anliker et al., 2020). Therefore, only the detected features which had decreasing intensity trends over time were significant for the assessment of the WWTP. The decrease in intensity demonstrated a higher load of features in the first studied samples (time points) of the last two months of 2018. During these months some productions resulted in higher pollution of wastewater compared to the remaining time studied.

Furthermore, these features of the WWTP influent (SP 3) prioritised in case of having increase/decrease in intensity over time, were assigned in the WWTP effluent, resulting in three features (cf. Appendix 5.6.2.2, Table-A 5-7). Two of them had decreasing intensity trends similar to the WWTP influent samples and one had no trend in intensity over time. Using these three features, the treatment process of the studied WWTP was assessed (see Figure 5-8).

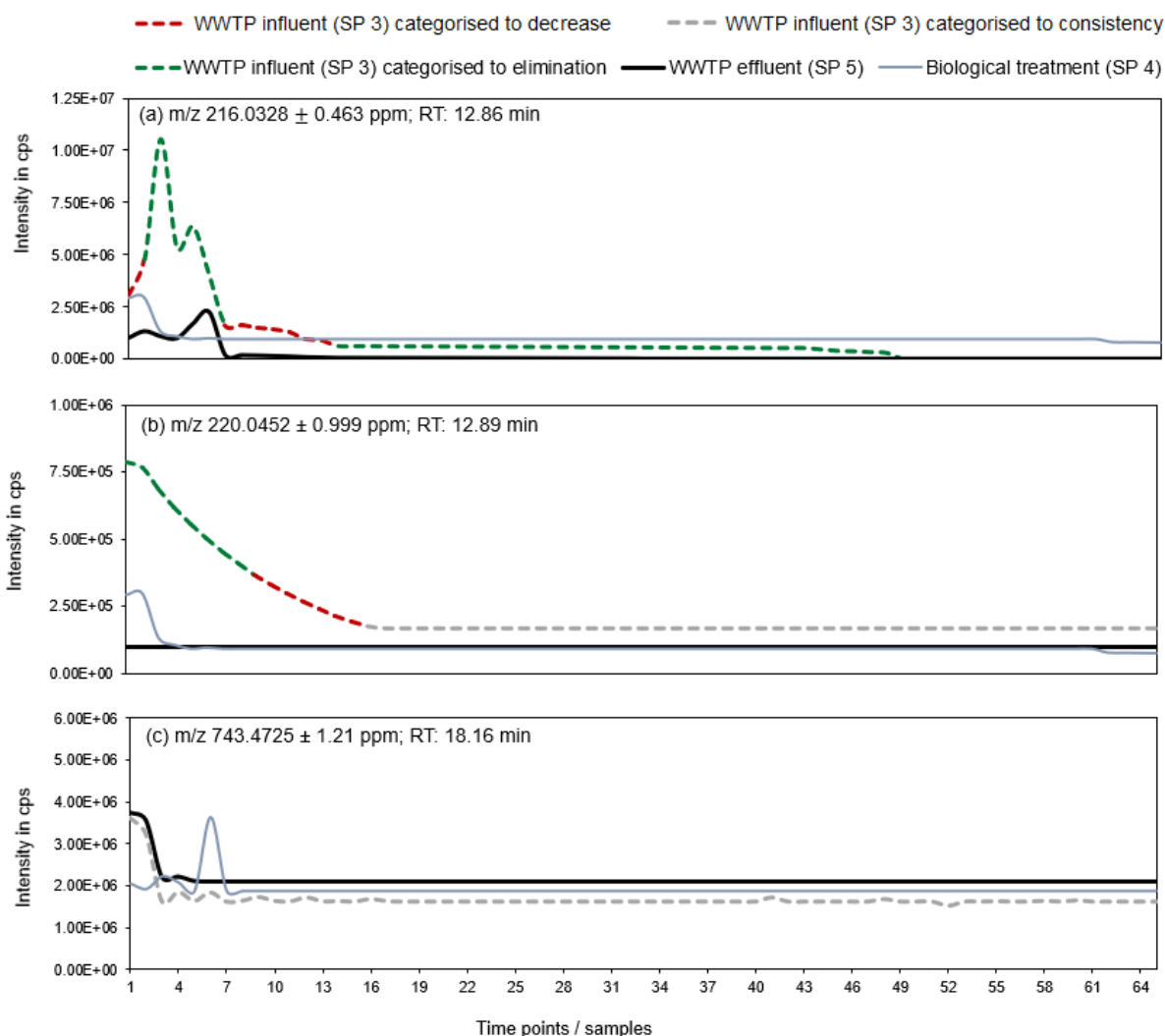


Figure 5-8 – Comparison of influent and effluent WWTP samples over time, demonstrating the treatment efficiency of the industrial WWTP. The three parts of the figure correspond to features prioritised for WWTP influent samples (SP 3) according to a decreasing trend in intensity over time which also were not fully eliminated during treatment processes (still detectable in SP 5). The given mass error refers to the accurate mass of aligned features of influent and effluent samples. The corresponding colour scheme illustrates the pairwise comparison of the process influent and effluent (SP 3 to SP 5) shown in influent data, resulting in categorisation of features into elimination, decrease or consistency based on the calculated fold change (Bader et al., 2017). For completeness, the features detection in sampling site SP 4 is shown, too.

Figure 5-8 a) shows the purification of the feature with the accurate mass of  $m/z$  216.0329 and a retention time of 12.8 min. The first significant peak of the third time point of the WWTP influent shows that the feature was almost completely degraded. There were only 10.2% of the intensity of the inflow detected in the WWTP effluent. However, the delay in flow time of the monitored WWTP should be considered. Depending on the weather of sampling time, the delay could be expanded to days (Harmon et al., 2011). Therefore, the first significant peak in time series of the WWTP effluent (time point six) could also correspond to both significant peaks of the WWTP influent (time points three and five). As the colour scheme demonstrated, the feature was continuously eliminated during treatment except for two phases of decreasing.



The feature was prioritised because of having a decreasing trend in intensity over time in the WWTP influent and still was detected in the WWTP effluent, which indicated substantial but no full degradation within the WWTP. Thus, features should be prioritised according to the time trend in the WWTP influent and the corresponding  $f_c$  during the treatment process to select features, which withstand the treatment process. Figure 5-8 b) shows another case. The studied feature was not wholly degraded, a baseload remained (57.7%). Furthermore, the feature was categorised into consistency from time point 15 on. In sampling site SP 4, the degradation had a delay in time. Figure 5-8 c) represented an example for a feature that did not degrade over time. The course of the feature intensities in the wastewater treatment plant flow corresponded to that of the wastewater treatment plant inlet (see Figure 5-8 c). In fact, this feature increased in intensity during treatment but was not categorised to 'increase' category since the increase was not significant. Therefore, the feature was categorised to consistency. In sampling site SP 4, there was a peaking trend observed which might indicate influences in the treatment process. However, matrix effects play a significant role when analysing and comparing influent and effluent samples of WWTPs (Schollée et al., 2015). As discussed by Nürenberg et al. (2015), a compensation of matrix effects by dilution was only applicable for the features with sufficiently high intensities. Thus, the assigned increase in intensity observed for feature with  $m/z$   $743.4725 \pm 1.21$  ppm, 18.16 min could also be a result of ion suppression by the WWTP influent.

The prioritised features having decreasing intensity trends over time in WWTP influent samples, partially withstand treatment process and thus were detected in the WWTP effluent as well, were tentatively identified. Features were tentatively identified as level 3 candidates (Schymanski et al., 2014). In Figure 5-9, the corresponding MS/MS spectra, retention times as well as predicted sum formulae of the three prioritised features are shown. For a probable structure, a library spectrum match or diagnostic evidence will be required, which could result in a confirmed structure by the comparison of a corresponding reference standard.

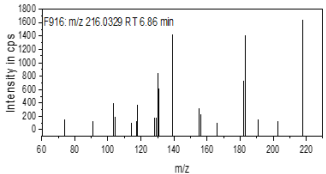
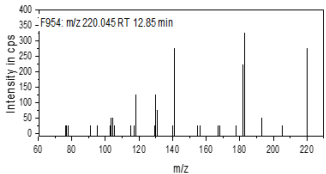
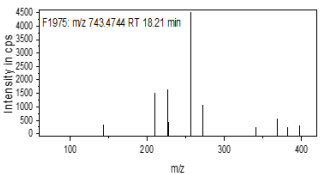
Accurate mass in Da	m/z 216.0329	m/z 220.0450	m/z 743.4744
Retention time in min	6.86	12.85	18.21
Ionisation	Positive	Positive	Positive
Proposed sum formulae	C <sub>4</sub> H <sub>7</sub> F <sub>2</sub> N <sub>3</sub> O <sub>5</sub>	C <sub>7</sub> H <sub>9</sub> NO <sub>7</sub>	C <sub>34</sub> H <sub>60</sub> F <sub>2</sub> N <sub>10</sub> O <sub>6</sub>
Score	95.6	88.2	97.3
Relative mass error in ppm	0.4	1.8	0.0
Hits in ChemSpider	0.0	4.0	0.0
MS/MS spectra			

Figure 5-9 – MS/MS-spectra of three detected peaks after treatment process out of temporal prioritised features of WWTP influent samples. Retention times (RT), accurate mass (m/z), as well as proposed sum formulae (based on ‘Non-targeted Screening’ workflow of SCIEX OS and ChemSpider data) with a relative mass deviation of the exact mass from the proposed sum formulae, are shown.

## 5.4 Conclusion

In this study, trend analyses based on spatial and temporal approaches were investigated to assess a WWTP. The studies of Bader et al. (2017) and Krauss et al. (2019) were expanded by time series trend analysis introduced in chapter 4 of this dissertation and applied to assess wastewater treatment processes. Thus, features with a significant site-specific contamination profile based on LC-HRMS data without any prior knowledge of the TrOCs present were proposed. Comprehensive chemical fingerprints of complex environmental mixtures were investigated, which strongly benefits the assessment of WWTPs. Features which were prioritised over time because of their relevant changes in intensity were all degraded during treatment with two exceptions. Furthermore, this study showed that treatment behaviour was almost constant over time by monitoring sampling sites during the treatment process over time. Altogether, this study demonstrated the usefulness of applying NTS in the assessment of the wastewater treatment process, which expanded the assessment to unknown compounds.

## 5.5 References

- Alygizakis, N.A., Gago-Ferrero, P., Hollender, J., Thomaidis, N.S., 2019. Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater. *J. Hazard. Mater.* 361, 19–29. <https://doi.org/10.1016/j.jhazmat.2018.08.073>
- Anliker, S., Loos, M., Comte, R., Ruff, M., Fenner, K., Singer, H., 2020. Assessing Emissions from Pharmaceutical Manufacturing Based on Temporal High-Resolution Mass Spectrometry Data. *Environ. Sci. Technol.* 54, 4110–4120. <https://doi.org/10.1021/acs.est.9b07085>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2017. LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment

- of Water Treatment Processes. *Anal. Chem.* 89, 13219–13226. <https://doi.org/10.1021/acs.analchem.7b03037>
- Barakat, A., El Baghdadi, M., Rais, J., Aghezzaf, B., Slassi, M., 2016. Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *Int. Soil Water Conserv. Res.* 4, 284–292. <https://doi.org/10.1016/j.iswcr.2016.11.002>
- Brack, W., Escher, B.I., Müller, E., Schmitt-Jansen, M., Schulze, T., Slobodnik, J., Hollert, H., 2018. Towards a holistic and solution-oriented monitoring of chemical status of European water bodies: how to support the EU strategy for a non-toxic environment? *Environ. Sci. Eur.* 30, 1–11. <https://doi.org/10.1186/s12302-018-0161-1>
- Buttiglieri, G., Peschka, M., Frömel, T., Müller, J., Malpei, F., Seel, P., Knepper, T.P., 2009. Environmental occurrence and degradation of the herbicide n-chloridazon. *Water Res.* 43, 2865–2873. <https://doi.org/10.1016/j.watres.2009.03.035>
- Camacho, J., Theron, R., Garcia-Gimenez, J.M., Macia-Fernandez, G., Garcia-Teodoro, P., 2019. Group-Wise Principal Component Analysis for Exploratory Intrusion Detection. *IEEE Access* 7, 113081–113093. <https://doi.org/10.1109/access.2019.2935154>
- Deeb, A.A., Stephan, S., Schmitz, O.J., Schmidt, T.C., 2017. Suspect screening of micropollutants and their transformation products in advanced wastewater treatment. *Sci. Total Environ.* 601–602, 1247–1253. <https://doi.org/10.1016/j.scitotenv.2017.05.271>
- Harmon, J., Theodore, M.G., Wardle, L.P., 2011. Methods to reduce wastewater treatment plant footprint and costs. 7,867,398 B2.
- Haun, J., Leonhardt, J., Portner, C., Hetzel, T., Tuerk, J., Teutenberg, T., Schmidt, T.C., 2013. Online and splitless NanoLC × CapillaryLC with quadrupole/time-of-flight mass spectrometric detection for comprehensive screening analysis of complex samples. *Anal. Chem.* 85, 10083–10090. <https://doi.org/10.1021/ac402002m>
- Hedgespeth, M.L., Gibson, N., Mccord, J., Shea, D., Nichols, E.G., 2019. Suspect screening and prioritization of chemicals of concern (COCs) in a forest-water reuse system watershed. *Sci. Total Environ.* 694. <https://doi.org/10.1016/j.scitotenv.2019.07.184>
- Hermes, N., Jewell, K.S., Wick, A., Ternes, T.A., 2018. Quantification of more than 150 micropollutants including transformation products in aqueous samples by liquid chromatography-tandem mass spectrometry using scheduled multiple reaction monitoring. *J. Chromatogr. A* 1531, 64–73. <https://doi.org/10.1016/j.chroma.2017.11.020>
- Hohrenk, L.L., Vosough, M., Schmidt, T.C., 2019. Implementation of Chemometric Tools to Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes. *Anal. Chem.* 91, 9213–9220. <https://doi.org/10.1021/acs.analchem.9b01984>
- Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* 184, 25–32. <https://doi.org/10.1016/j.envpol.2013.07.048>
- Huntscha, S., Hofstetter, T.B., Schymanski, E.L., Spahr, S., Hollender, J., 2014. Biotransformation of benzotriazoles: Insights from transformation product identification and compound-specific isotope analysis. *Environ. Sci. Technol.* 48, 4435–4443. <https://doi.org/10.1021/es405694z>
- Itzel, F., Baetz, N., Hohrenk, L.L., Gehrman, L., Antakyali, D., Schmidt, T.C., Tuerk, J., 2020. Evaluation of a biological post-treatment after full-scale ozonation at a

- municipal wastewater treatment plant. *Water Res.* 170, 115316. <https://doi.org/10.1016/j.watres.2019.115316>
- Jobst, K.J., Shen, L., Reiner, E.J., Taguchi, V.Y., Helm, P.A., McCrindle, R., Backus, S., 2013. The use of mass defect plots for the identification of (novel) halogenated contaminants in the environment. *Anal. Bioanal. Chem.* 405, 3289–3297. <https://doi.org/10.1007/s00216-013-6735-2>
- Kiefer, K., Müller, A., Singer, H., Hollender, J., 2019. New relevant pesticide transformation products in groundwater detected using target and suspect screening for agricultural and urban micropollutants with LC-HRMS. *Water Res.* 165, 114972. <https://doi.org/10.1016/j.watres.2019.114972>
- Krauss, M., Hug, C., Bloch, R., Schulze, T., Brack, W., 2019. Prioritising site-specific micropollutants in surface water from LC-HRMS non-target screening data using a rarity score. *Environ. Sci. Eur.* 31, 45. <https://doi.org/10.1186/s12302-019-0231-z>
- Leendert, V., Van Langenhove, H., Demeestere, K., 2015. Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments. *TrAC - Trends Anal. Chem.* 67, 192–208. <https://doi.org/10.1016/j.trac.2015.01.010>
- Loos, M., Singer, H., 2017. Nontargeted homologue series extraction from hyphenated high resolution mass spectrometry data. *J. Cheminform.* 9, 1–11. <https://doi.org/10.1186/s13321-017-0197-z>
- Nürnberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on LC-HRMS analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>
- Özkaraoğlu, E.B., Akbal, F., Kuleyin, A., 2018. Potential reuse of treated industrial wastewater in agriculture: Textile wastewater, in: *International Scientific Journal Mechanisation in Agriculture*. pp. 138–140.
- Park, M., Snyder, S.A., 2020. Chemosphere Statistical profiling for identifying transformation products in an engineered treatment process. *Chemosphere* 251. <https://doi.org/10.1016/j.chemosphere.2020.126401>
- Parry, E., Young, T.M., 2016. Comparing targeted and non-targeted high-resolution mass spectrometric approaches for assessing advanced oxidation reactor performance. *Water Res.* 104, 72–81. <https://doi.org/10.1016/j.watres.2016.07.056>
- Ranade, V. V., Bhandari, V.M., 2014. *Industrial Wastewater Treatment, Recycling, and Reuse-Past, Present and Future, Industrial Wastewater Treatment, Recycling and Reuse*. Elsevier Ltd. <https://doi.org/10.1016/B978-0-08-099968-5.00014-3>
- Richardson, S.D., Kimura, S.Y., 2016. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* 88, 546–582. <https://doi.org/10.1021/acs.analchem.5b04493>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Ruppe, S., Griesshaber, D.S., Langlois, I., Singer, H.P., Mazacek, J., 2018. Detective Work on the Rhine River in Basel – Finding Pollutants and Polluters. *Chim. Int. J. Chem.* 72, 547. <https://doi.org/10.2533/chimia.2018.547>
- Schlüsener, M.P., Kunkel, U., Ternes, T.A., 2015. Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants. *Environ. Sci. Technol.* 49, 14282–14291. <https://doi.org/10.1021/acs.est.5b03926>

- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* 87, 12121–12129. <https://doi.org/10.1021/acs.analchem.5b02905>
- Schollée, J.E., Schymanski, E.L., Hollender, J., 2016. Statistical Approaches for LC-HRMS Data to Characterize, Prioritize, and Identify Transformation Products from Water Treatment Processes. *ACS Symp. Ser.* 1241, 45–65. <https://doi.org/10.1021/bk-2016-1241.ch004>
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. <https://doi.org/10.1021/es5002105>
- Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - A case study. *Water Res.* 38, 3980–3992. <https://doi.org/10.1016/j.watres.2004.06.011>
- Verkh, Y., Rozman, M., Petrovic, M., 2018. A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. *Chemosphere* 200, 397–404. <https://doi.org/10.1016/j.chemosphere.2018.02.095>
- Verlicchi, P., Al Aukidy, M., Zambello, E., 2012. Occurrence of pharmaceutical compounds in urban wastewater: Removal, mass load and environmental risk after a secondary treatment-A review. *Sci. Total Environ.* 429, 123–155. <https://doi.org/10.1016/j.scitotenv.2012.04.028>
- Yan, S., Wang, M., Zha, J., Zhu, L., Li, W., Luo, Q., Sun, J., Wang, Z., 2017. Environmentally relevant concentrations of carbamazepine caused endocrine-disrupting effects on non-target organisms, Chinese rare minnows (*Gobiocypris rarus*). *Environ. Sci. Technol.* 52, 886–894. <https://doi.org/10.1021/acs.est.7b06476>
- Zeinalzadeh, K., Rezaei, E., 2017. Determining spatial and temporal changes of surface water quality using principal component analysis. *J. Hydrol. Reg. Stud.* 13, 1–10. <https://doi.org/10.1016/j.ejrh.2017.07.002>

## 5.6 Chapter Appendix

### 5.6.1 Supplement for Chapter Materials and Method

#### 5.6.1.1 *Used Target Analytes and Internal Standards*

Table-A 5-1 – Targets and ISTD used for testing data set covering relevant and non-relevant time trends (\*industrial chemical).

Target compounds	Sum formula	CAS No.	Ionisation	RT in min	m/z [ESI*]	m/z [ESI*]	Conc. Range in µg/L
2-(Trifluoromethyl)benzamide	C <sub>8</sub> H <sub>6</sub> F <sub>3</sub> NO	360-64-5	[M+H] <sup>+</sup>	5.32	190.05	-	0.200 - 30.0
Acifluorfen	C <sub>14</sub> H <sub>7</sub> ClF <sub>3</sub> NO <sub>5</sub>	50594-66-5	[M-H] <sup>-</sup>	13.77	-	359.99	2.00 - 100
Azoxystrobin	C <sub>22</sub> H <sub>17</sub> N <sub>3</sub> O <sub>5</sub>	131860-33-8	[M+H] <sup>+</sup>	13.07	404.12	-	0.200 - 200
Azoxystrobin-D4	C <sub>22</sub> H <sub>13</sub> D <sub>4</sub> N <sub>3</sub> O <sub>5</sub>	-	[M+H] <sup>+</sup>	13.04	408.15	-	0.0250
ABC 700*	C <sub>8</sub> H <sub>6</sub> F <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	-	[M+H] <sup>+</sup>	3.56	177.05	-	2.00 - 130
Bezafibrate-D6	C <sub>19</sub> H <sub>14</sub> D <sub>6</sub> CINO <sub>4</sub>	1219802-74-0	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	12.62 / 12.61	368.15	366.14	25.0
Carbamazepine	C <sub>15</sub> H <sub>12</sub> N <sub>2</sub> O	298-46-4	[M+H] <sup>+</sup>	10.84	237.1	-	0.200 - 200
Chloramphenicol-D5	C <sub>11</sub> H <sub>7</sub> D <sub>5</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>5</sub>	202480-68-0	[M-H] <sup>-</sup>	6.12	-	326.04	25.0
Climbazole	C <sub>15</sub> H <sub>17</sub> ClN <sub>2</sub> O <sub>2</sub>	38083-17-9	[M+H] <sup>+</sup>	10.14	293.11	-	0.200 - 17.0
Clothianidin	C <sub>6</sub> H <sub>8</sub> ClN <sub>5</sub> O <sub>2</sub> S	2180880-92-5	[M+H] <sup>+</sup>	6.15	250.02	-	10.0 - 13.0
Clothianidin-D3	C <sub>6</sub> H <sub>5</sub> D <sub>3</sub> ClN <sub>5</sub> O <sub>2</sub> S	1262776-24-8	[M+H] <sup>+</sup>	6.12	253.04	-	25.0
Cyproconazole	C <sub>15</sub> H <sub>18</sub> ClN <sub>3</sub> O	94361-06-5	[M+H] <sup>+</sup>	13.73	292.12	-	0.200 - 200
Diclofenac-D4	C <sub>14</sub> H <sub>7</sub> D <sub>4</sub> NO <sub>2</sub> Cl <sub>2</sub>	153466-65-0	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	14.11	300.05	298.04	25.0
Diethylene glycol dimethyl ether/ Diglyme	C <sub>6</sub> H <sub>14</sub> O <sub>3</sub>	111-96-6	[M+H] <sup>+</sup>	4.62	135.1	-	0.200 - 200
Diphenyl sulfone	C <sub>12</sub> H <sub>10</sub> O <sub>2</sub> S	127-63-9	[M+NH <sub>4</sub> ] <sup>+</sup>	10.15	236.07	-	0.200 - 17.0
Diuron-D6	C <sub>9</sub> H <sub>4</sub> D <sub>6</sub> Cl <sub>2</sub> N <sub>2</sub> O	1007536-67-5	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	11.95 / 11.93	239.06	237.05	25.0
Lactofen	C <sub>19</sub> H <sub>15</sub> ClF <sub>3</sub> NO <sub>7</sub>	77501-63-4	[M+NH <sub>4</sub> ] <sup>+</sup>	16.53	479.08	-	0.200 - 35.0
Mecoprop-D3	C <sub>10</sub> D <sub>3</sub> H <sub>8</sub> ClO <sub>3</sub>	352431-15-3	[M-H] <sup>-</sup>	12.23	-	216.05	25.0
Mesotrione-D4	C <sub>14</sub> H <sub>9</sub> D <sub>4</sub> NO <sub>7</sub> S	-	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	8.52 / 8.53	344.07	342.06	25.0
Metsulfuron-methyl-D3	C <sub>14</sub> H <sub>12</sub> D <sub>3</sub> N <sub>5</sub> O <sub>6</sub> S	-	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	10.18 / 10.17	385.1	383.09	25.0
N,N-Diethyl-3-methylbenzamide/ DEET	C <sub>12</sub> H <sub>17</sub> NO	134-62-3	[M+H] <sup>+</sup>	11.67	192.14	-	2.00 - 100
N,N-Diethyl-3-methyl-D3-benzamide-D4/ DEET-D7	C <sub>12</sub> H <sub>10</sub> D <sub>7</sub> NO	1219799-37-7	[M+H] <sup>+</sup>	11.6	199.18	-	1.25
N,N-Dimethyl-N'-phenylsulfamid/ DMSA	C <sub>8</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub> S	4710-17-2	[M+H] <sup>+</sup>	7.42	201.07	-	0.200 - 17.0
Prothioconazole	C <sub>14</sub> H <sub>15</sub> Cl <sub>2</sub> N <sub>3</sub> O <sub>5</sub>	178928-70-6	[M+H] <sup>+</sup>	14.74	344.04	-	20.0 - 200
Salicylic acid-D4	C <sub>7</sub> H <sub>2</sub> D <sub>4</sub> O <sub>3</sub>	78646-17-0	[M-H] <sup>-</sup>	6.22	-	141.05	25.0
Sulcotrione-D4	C <sub>14</sub> D <sub>4</sub> H <sub>9</sub> ClO <sub>5</sub> S	-	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	9.71 / 9.70	333.05	331.04	0.250
Tebuconazole	C <sub>16</sub> H <sub>22</sub> ClN <sub>3</sub> O	107534-96-3	[M+H] <sup>+</sup>	14.94	308.15	-	0.200 - 17.0
Tebuconazole-D6	C <sub>16</sub> H <sub>16</sub> D <sub>6</sub> ClN <sub>3</sub> O	-	[M+H] <sup>+</sup>	14.91	314.19	-	1.25
Tetraethyleneglycoldimethylether/ Tetraglyme	C <sub>10</sub> H <sub>22</sub> O <sub>5</sub>	143-24-8	[M+H] <sup>+</sup>	6.82	223.15	-	0.200 - 17.0
Thiacloprid	C <sub>10</sub> H <sub>9</sub> ClN <sub>4</sub> S	111988-49-9	[M+H] <sup>+</sup>	8.63	253.03	-	0.200 - 17.0
Triethyleneglycoldimethylether/ Triglyme	C <sub>8</sub> H <sub>18</sub> O <sub>4</sub>	112-49-2	[M+H] <sup>+</sup>	5.8	179.13	-	0.200 - 200

5.6.1.2 LC-HRMS Method Parameters

Table-A 5-2 – LC-HRMS acquisition method.

<b>Parameter</b>	<b>Setting</b>
LC system	Agilent 1290 Infinity
LC column	Raptor™ Ultra Aqueous C18 analytical column  (100 mm x 2.1 mm, 3.0 µm particle size)
Mobile Phase A	0.1% formic acid in Milli-Q®
Mobile Phase B	0.1% formic acid in methanol
Gradient (%B)	0.5 min (0% B), 1.0 min (10% B), 20 min (90% B), 26 min (90% B), 32 min (0% B)
Flow rate	300 µL min <sup>-1</sup>
Temperature	55 °C
Sample injection volume	5 µL
MS system	x500R qTOF (SCIEX)
Ion source	Electrospray ion source Turbo V™
MS mass range	<i>m/z</i> 70 – 950
Ion source gas 1	50
Ion source gas 2	70
Curtain gas	40
Source temperature	450 °C
IonSpray voltage floating	5500 V / –4500 V
Declustering potential	80 V
Collision energy	7 V
MS/MS Experiment	10 data-dependent TOF-MS/MS scans
MS/MS mass range	<i>m/z</i> 30 – 700

### 5.6.1.3 Data Treatment using MarkerView

Table-A 5-3 – MarkerView was used for peak detection, peak alignment, normalisation of retention times and blank value subtraction. The following criteria were used. The average values of peak intensities and retention times of internal standards (ISTD, see 5.6.1.1) were used to correct the retention times (RT) and to normalise the peak intensities.

<b>Parameters</b>	<b>Settings</b>
Import data type	LC/MS data from .wiff2 files
<b>Peak Finding Options</b>	
Minimum RT	1 min
Maximum RT	26 min
Subtraction offset	10 scans
Minimum spectral peak width	30 ppm
Subtraction mult. Factor	2
Minimum RT peak width	8 scans
Noise threshold	1000
Assign Charge States	Yes
<b>Alignment &amp; Filtering</b>	
RT tolerance	2%
Mass tolerance	3 ppm
Maximum number of peaks	10 <sup>9</sup>
Isotope filtering	Remove isotopes (keeping peaks with unknown status)
RT correction using ISTD	Correction type: linear, RT tolerance: 10%, mass tolerance: 2 ppm
Sample normalisation using ISTD	RT tolerance: 10%, mass tolerance: 2 ppm



5.6.1.4 Quality Control of ISTD during Data Analysis

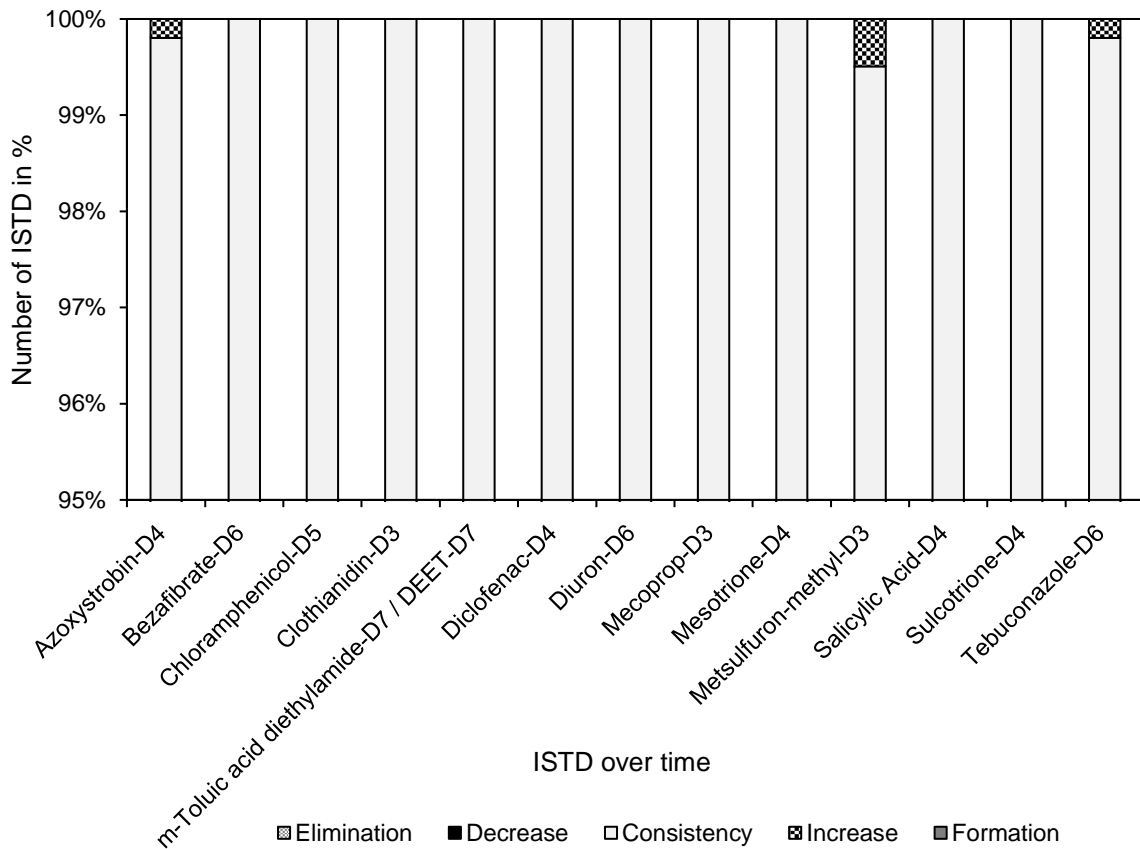


Figure-A 5-1 – Result of quality control during data analysis using the evaluation scheme of Bader et al. (2017). The comparison of WWTP effluent and influent of samples taken over five months is shown focussing on spiked ISTD. Evaluation is based on Bader et al. (2017). The quotient of SP 5/SP 3 was built by dividing every day's intensity from another.

## 5.6.2 Supplement for Chapter Results

### 5.6.2.1 Analysis of Sampling Sites

Table-A 5-4 - Site-specific features of sampling stations SP 1 to SP 5 The mean intensity was averaged over all sampling times. RT  $\triangleq$  retention time.

SP 1				SP 2			
Accurate mass in Da	RT in min	Trend	Mean intensity in cps over time	Accurate mass in Da	RT in min	Trend	Mean intensity in cps over time
87.0032	3.31	no trend	2.45E+05	70.0649	2.54	no trend	9.27E+04
87.0035	1.14	no trend	8.67E+04	70.065	2.98	no trend	1.27E+05
144.0305	14.12	no trend	6.88E+04	70.065	3.13	no trend	8.76E+04
181.9916	4.05	peaks	4.56E+04	71.0131	2.01	no trend	1.81E+05
225.0629	2.72	no trend	6.95E+04	71.0132	1.10	no trend	1.05E+05
226.1416	8.89	no trend	1.09E+05	71.0132	1.87	no trend	1.12E+05
252.1976	17.83	no trend	1.26E+05	74.0962	3.04	no trend	1.19E+05
257.2107	18.51	no trend	2.56E+05	75.0089	1.33	no trend	9.87E+04
266.1031	16.35	no trend	9.63E+04	76.0392	2.82	no trend	3.75E+04
298.2733	17.31	no trend	3.48E+05	76.0392	3.73	no trend	3.30E+04
304.2483	16.26	no trend	2.15E+05	93.0573	4.39	no trend	3.35E+05
331.1882	16.14	no trend	1.27E+05	114.1276	5.27	no trend	2.13E+05
345.3019	17.61	no trend	1.39E+05	119.0488	9.15	no trend	5.87E+04
350.2299	17.88	peaks	4.30E+05	125.9866	5.33	no trend	4.18E+04
353.2678	19.58	no trend	1.64E+05	140.0816	5.80	no trend	6.84E+04
				158.1537	11.96	no trend	2.13E+05
				163.9847	12.13	no trend	5.61E+04
				171.044	13.11	no trend	1.08E+05
				172.0154	2.72	peaks	8.09E+04
				180.1387	5.94	no trend	1.89E+05
				190.1358	3.80	no trend	1.23E+05
				206.9697	7.55	no trend	1.13E+05
				211.0949	14.05	no trend	1.35E+05
				222.1702	2.83	no trend	3.02E+05
				233.211	15.11	no trend	1.80E+05
				269.2087	17.44	no trend	3.32E+05
				282.2407	19.86	no trend	3.56E+04
				372.3834	17.16	no trend	4.96E+05
				381.1375	16.78	no trend	1.87E+05
				390.7754	15.58	no trend	4.71E+05
				497.8482	16.13	no trend	1.39E+05
				708.5071	23.21	no trend	8.51E+05
				948.7187	21.35	no trend	1.02E+06

SP 3				SP 4			
Accurate mass in Da	RT in min	Trend	Mean intensity in cps over time	Accurate mass in Da	RT in min	Trend	Mean intensity in cps over time
71.014	1.72	no trend	1.58E+05	71.0603	13.00	no trend	3.76E+04
78.9942	3.38	no trend	7.27E+04	72.0091	9.53	no trend	3.58E+04
129.0195	3.87	no trend	4.05E+04	72.0091	10.03	no trend	3.35E+04
172.0978	5.84	no trend	3.42E+04	72.0805	2.97	no trend	1.48E+05
172.8428	1.14	no trend	4.46E+04	72.0806	3.87	no trend	5.84E+04
191.0196	1.71	no trend	1.77E+05	72.0806	4.05	no trend	6.84E+04
228.1962	14.61	single peak	1.64E+05	72.0806	4.17	no trend	7.99E+04
243.2792	12.08	no trend	9.55E+05	79.0209	3.45	no trend	1.92E+05
244.1545	10.56	no trend	2.87E+05	86.0459	2.86	no trend	8.66E+04
249.1098	10.56	no trend	7.45E+05	117.0698	6.36	no trend	1.04E+05
264.2325	11.56	single peak	1.46E+05	121.0648	9.82	no trend	6.08E+04
271.1537	11.31	no trend	1.19E+06	123.0915	4.53	no trend	1.24E+05
272.0957	6.59	no trend	1.13E+06	129.0558	5.59	no trend	7.62E+04

Chapter 5 Spatial Trend detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

Table-A 5-4 continued.

274.1109	6.17	no trend	3.43E+05	130.0432	7.73	no trend	5.97E+04
277.1279	11.12	peaks	1.77E+05	132.9632	4.01	peaks	5.77E+04
296.1283	16.19	no trend	2.37E+05	134.0711	9.24	no trend	2.12E+05
359.2066	12.35	no trend	5.67E+05	141.0105	3.71	peaks	7.39E+04
375.0108	5.27	no trend	1.55E+05	142.0416	5.60	single peak	7.92E+04
377.0256	5.26	no trend	1.58E+05	143.0815	5.67	peaks	1.52E+05
381.1885	12.35	no trend	7.04E+05	145.0096	5.38	no trend	6.79E+04
675.5006	20.89	no trend	4.72E+05	147.9999	6.07	no trend	5.59E+04
				149.0009	2.11	single peak	2.43E+05
				150.1044	4.57	no trend	1.29E+05
				151.0309	13.02	increase	6.99E+04
				156.0124	12.61	no trend	7.28E+04
				160.0868	11.79	no trend	6.65E+04
				163.0636	8.04	single peak	4.29E+05
				167.1066	11.44	no trend	5.42E+04
				170.0965	9.96	no trend	4.39E+04
				171.1488	3.68	single peak	9.35E+04
				177.0912	8.84	no trend	1.29E+05
				191.0371	13.78	no trend	7.99E+05
				200.9021	7.92	no trend	8.51E+04
				223.0521	13.81	no trend	9.05E+04
				228.0701	10.86	no trend	2.90E+05
				229.1042	8.43	no trend	1.31E+05
				231.1602	16.05	no trend	1.78E+05
				232.0085	4.78	no trend	2.47E+05
				242.0853	12.33	no trend	9.45E+04
				243.1356	15.78	no trend	6.01E+04
				262.111	9.76	increase	1.21E+05
				263.0745	13.48	no trend	6.49E+04
				290.9625	15.49	no trend	1.43E+05
				303.0149	13.32	no trend	1.12E+05
				324.2012	5.49	no trend	5.20E+05
				338.9865	14.04	single peak	1.68E+05
				535.2699	16.90	no trend	5.01E+05

<b>SP 5</b>			
<b>Accurate mass in Da</b>	<b>RT in min</b>	<b>Trend</b>	<b>Mean intensity in cps over time</b>
103.0391	9.36	no trend	4.21E+04
120.0557	6.52	single peak	5.71E+04
142.0693	4.34	no trend	9.83E+04
168.1161	20.54	no trend	9.42E+04
228.1487	17.87	no trend	9.23E+04
232.108	6.45	single peak	1.08E+05
233.0312	12.77	no trend	1.00E+05
298.0341	13.93	no trend	1.27E+05
531.1001	5.42	no trend	3.10E+05

Table-A 5-5 – Prioritised features by rarity score introduced by Krauss et al. in 2019. RT  $\hat{=}$  retention time.

Accurate mass in Da	Retention time in min	Ionisation	Maximum intensity across all sites	Median intensity across all sites	Positive defects across sampling sites	Rarity score	Purposed sum formulae	Score of sum formulae	Relative mass error in ppm
m/z 125.1072	7.23	positive	1.28E+06	3.05E+04	SP 4	105	C <sub>6</sub> H <sub>8</sub> N <sub>2</sub> O	78.1	0.5
m/z 142.0693	4.34	positive	1.18E+06	2.46E+04	SP 2	120	C <sub>7</sub> H <sub>11</sub> NS	63.7	6.4
m/z 144.9648	1.92	positive	3.72E+06	8.29E+04	SP 2	112	CHFO <sub>7</sub>	94.1	0.6
m/z 145.0158	4.27	positive	1.54E+06	5.90E+04	SP 1	131	C <sub>5</sub> H <sub>5</sub> ClN <sub>2</sub> O	69.8	3.0
m/z 150.1276	8.28	positive	3.94E+06	7.29E+04	SP 2	135	C <sub>6</sub> H <sub>15</sub> NO <sub>3</sub>	87.0	0.6
m/z 151.0614	4.17	positive	4.77E+06	2.87E+04	SP 4, SP 3	416	C <sub>6</sub> H <sub>6</sub> N <sub>4</sub> O	84.5	0.6
m/z 157.0771	8.17	positive	2.15E+06	3.88E+04	SP 4	139	C <sub>7</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	80.6	1.9
m/z 171.0662	3.41	positive	1.07E+06	3.90E+04	SP 4	137	CH <sub>11</sub> N <sub>6</sub> O <sub>2</sub> P	73.9	3.9
m/z 204.1131	5.14	positive	1.89E+07	8.58E+05	SP 4	110	C <sub>12</sub> H <sub>13</sub> NO <sub>2</sub>	89.1	1.1
m/z 206.1176	11.09	positive	4.65E+06	5.41E+04	SP 5, SP 4	215	No formula found	-	-
m/z 207.006	3.31	negative	7.43E+06	6.02E+04	SP 3, SP 2	309	CH <sub>5</sub> N <sub>8</sub> OPS	96.6	0.3
m/z 213.0226	6.69	positive	1.18E+06	3.65E+04	SP 4	162	C <sub>2</sub> F <sub>4</sub> N <sub>8</sub>	85.8	2.1
m/z 215.8871	2.67	positive	2.40E+07	6.14E+05	SP 4	195	C <sub>3</sub> Cl <sub>2</sub> NO <sub>2</sub> PS	73.2	8.0
m/z 218.0303	14.82	positive	3.41E+06	4.36E+04	SP 3, SP 4	195	C <sub>6</sub> H <sub>13</sub> F <sub>2</sub> NO <sub>2</sub> S <sub>2</sub>	87.1	0.7
m/z 219.9836	6.55	positive	1.77E+06	4.97E+04	SP 4	178	C <sub>2</sub> H <sub>6</sub> FN <sub>3</sub> O <sub>3</sub> S <sub>2</sub>	99.7	0.0
m/z 220.045	12.85	positive	8.31E+06	1.05E+05	SP 2	198	C <sub>2</sub> H <sub>6</sub> N <sub>9</sub> O <sub>2</sub> P	91.9	0.3
m/z 224.1644	5.53	positive	3.51E+06	7.57E+04	SP 5	116	C <sub>7</sub> H <sub>22</sub> N <sub>5</sub> OP	55.4	6.7
m/z 234.0867	7.16	positive	3.90E+06	7.52E+04	SP 4	259	C <sub>2</sub> H <sub>13</sub> FN <sub>7</sub> O <sub>3</sub> P	77.7	2.5
m/z 237.0863	5	positive	7.77E+06	3.26E+05	SP 3	119	C <sub>5</sub> H <sub>13</sub> N <sub>6</sub> O <sub>3</sub> P	79.3	1.1
m/z 242.211	19.8	positive	4.48E+06	7.04E+04	SP 2	159	C <sub>14</sub> H <sub>27</sub> NO <sub>2</sub>	37.4	7.0
m/z 251.0664	11.38	positive	3.07E+06	1.07E+05	SP 5	144	C <sub>12</sub> H <sub>3</sub> FN <sub>6</sub>	91.4	0.2
m/z 269.0237	6.06	positive	8.31E+06	1.08E+05	SP 3	192	CH <sub>10</sub> F <sub>2</sub> N <sub>8</sub> S <sub>3</sub>	92.3	0.2
m/z 269.9034	7.9	positive	8.31E+06	1.12E+05	SP 4	185	C <sub>3</sub> F <sub>4</sub> NO <sub>3</sub> PS <sub>3</sub>	69.6	3.0
m/z 289.0309	14.51	positive	3.47E+06	8.57E+04	SP 4	101	C <sub>4</sub> H <sub>10</sub> FN <sub>6</sub> O <sub>6</sub> P	98.6	0.1
m/z 295.9888	9.95	positive	4.24E+06	9.82E+04	SP 2	108	C <sub>4</sub> H <sub>9</sub> N <sub>9</sub> OS <sub>3</sub>	89.2	1.6
m/z 299.1319	10.97	negative	1.84E+06	8.30E+04	SP 4	111	C <sub>13</sub> H <sub>26</sub> F <sub>2</sub> OS <sub>2</sub>	93.9	0.7
m/z 309.0997	8.41	positive	3.31E+06	5.96E+04	SP 5	139	C <sub>14</sub> H <sub>20</sub> F <sub>3</sub> O <sub>2</sub> P	94.2	0.5
m/z 326.0503	15.38	negative	2.35E+06	5.12E+04	SP 4	115	C <sub>11</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub> S	100.0	0.0
m/z 337.8714	5.06	positive	2.66E+06	8.78E+04	SP 4	152	CH <sub>5</sub> FN <sub>10</sub> PS <sub>3</sub>	98.9	0.1
m/z 341.0987	8.95	negative	1.16E+07	2.68E+05	SP 4	109	C <sub>20</sub> H <sub>13</sub> F <sub>3</sub> O <sub>2</sub>	95.3	0.7
m/z 352.2328	4.34	positive	2.95E+06	1.25E+05	SP 4	118	C <sub>21</sub> H <sub>29</sub> N <sub>5</sub>	99.9	0.0
m/z 363.0809	7.57	negative	2.69E+06	7.94E+04	SP 5	169	C <sub>12</sub> H <sub>9</sub> FN <sub>8</sub> O <sub>5</sub>	94.6	0.2
m/z 435.3066	16.48	positive	1.59E+07	1.65E+05	SP 3, SP 2	240	C <sub>23</sub> H <sub>38</sub> N <sub>4</sub> O <sub>4</sub>	97.6	0.2
m/z 613.5516	20.08	positive	9.03E+07	5.66E+05	SP 1, SP 2	399	C <sub>36</sub> H <sub>70</sub> F <sub>2</sub> N <sub>4</sub> O	84.7	0.1
m/z 737.5655	19.77	positive	6.34E+06	7.23E+04	SP 1	219	No formula found	-	-

5.6.2.2 Spatial Trend Analysis

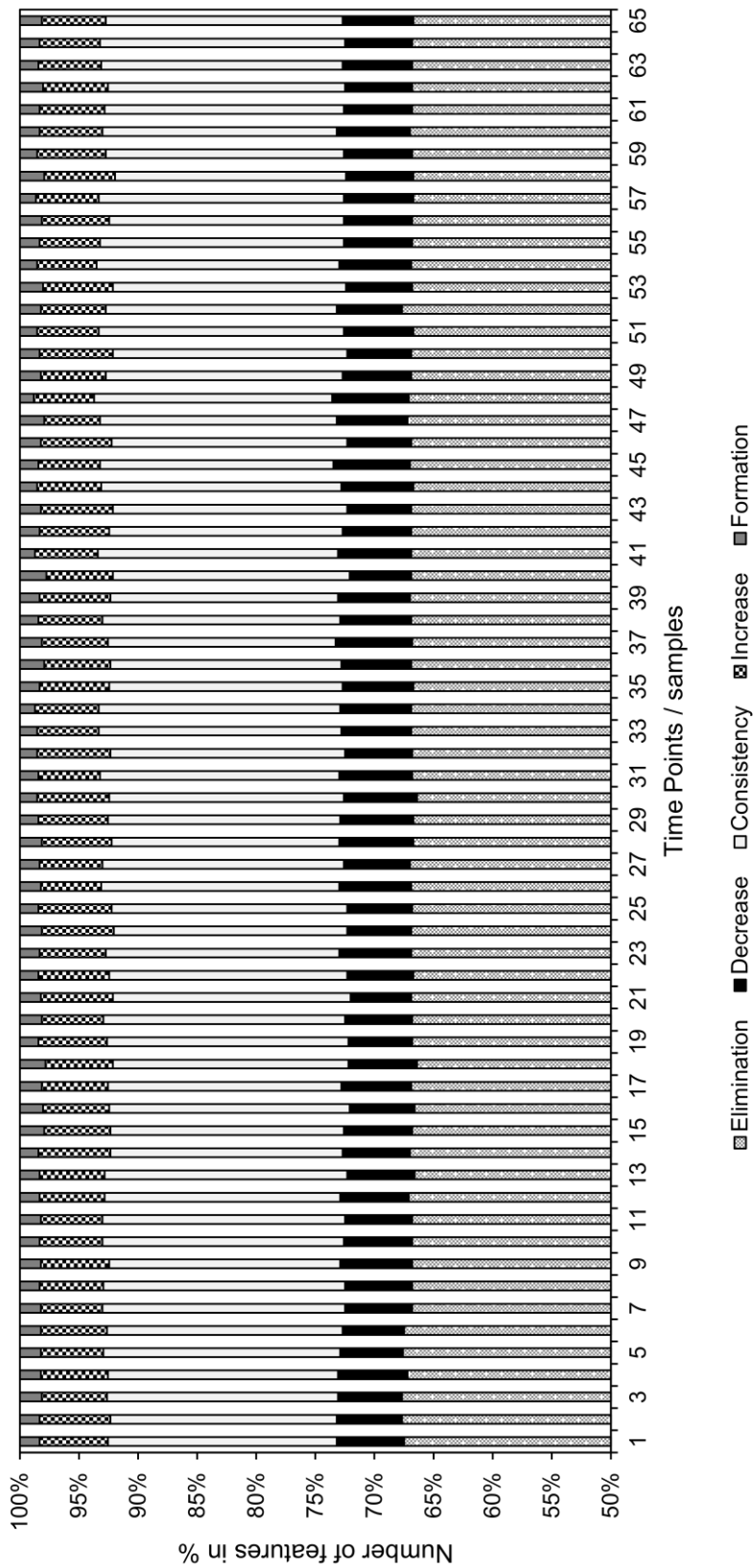


Figure-A 5-2 – Result of treatment efficiency of the WWTP. The pairwise comparison of process influent and effluent samples taken over five months, resulted in features assigned to the categories elimination, decrease, consistency, increase, or formation samples. The categorisation was calculated by the fold change (fc) approach introduced by Bader et al. (2017).

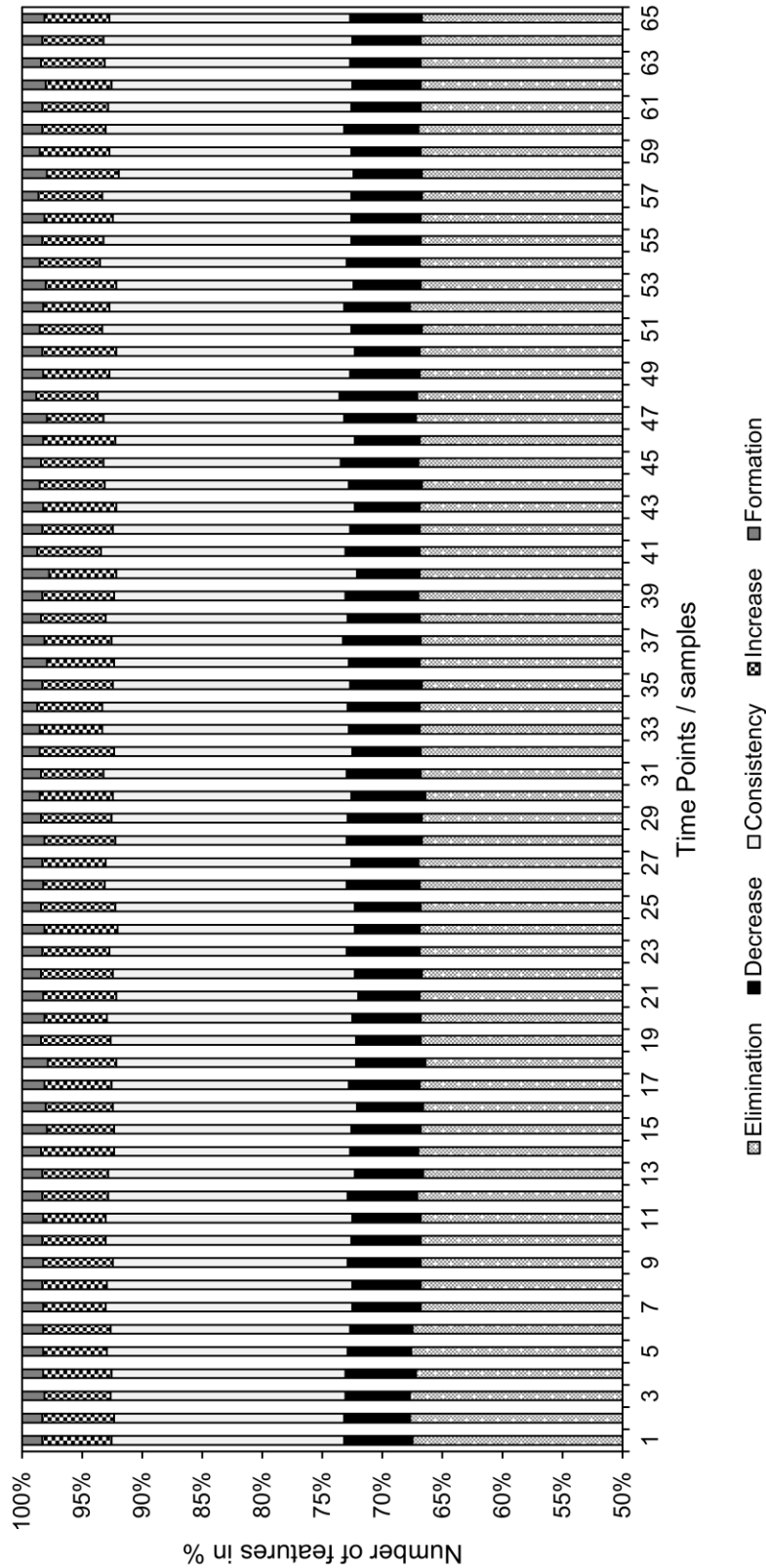


Figure-A 5-3 – Result of treatment efficiency of the WWTP considering the flow time in the WWTP. The assessment scheme of pairwise comparison introduced by Bader et al. (2017) was modified by calculating the fold change (fc) between the intensities of process influent sample of the studied day and the intensities of process effluent of the next day to check whether some features had a delay in flow time in the WWTP.

5.6.2.3 Temporal Trend Analysis

Table-A 5-6 – Features which were identified in both the WWTP influent (SP 3) and treated wastewater (SP 5) which consequently partially withstand treatment process. The m/z, retention time (RT), median intensity over all measured wastewater samples, as well as the corresponding trend are listed.

Feature m/z	Feature RT in min	Median intensity over all measured samples in cps	Trend
84.0553	3.28	2.09E+02	peaks
96.9595	1.13	8.08E+02	peaks
109.0754	3.32	4.56E+02	peaks
109.0756	4.00	5.77E+02	peaks
109.0757	3.76	3.69E+02	peaks
113.1068	2.43	3.63E+02	peaks
127.0723	1.32	2.88E+02	peaks
131.0424	5.01	8.73E+01	peaks
135.0261	4.17	1.06E+02	peaks
137.0344	5.01	1.27E+02	peaks
141.0653	4.33	7.62E+02	peaks
141.0654	4.94	6.15E+02	peaks
144.9647	1.30	1.48E+02	peaks
145.0160	4.98	2.57E+02	peaks
145.0160	4.42	2.42E+02	peaks
147.0133	4.96	1.14E+02	peaks
150.1273	8.60	7.09E+01	peaks
152.1178	6.73	1.65E+02	peaks
152.1179	7.23	1.37E+02	peaks
153.0908	17.95	1.39E+02	decrease
155.1065	17.97	2.05E+02	decrease
159.0910	6.03	2.79E+02	peaks
159.0912	4.92	6.00E+02	decrease
159.0913	6.53	4.05E+02	peaks
160.8424	1.11	2.14E+02	peaks
162.8393	1.11	1.84E+02	peaks
170.0726	4.47	6.25E+01	peaks
170.1285	6.80	1.39E+02	peaks
172.1328	17.98	1.72E+03	decrease
175.0316	5.10	1.86E+02	peaks
175.0322	5.03	3.99E+02	peaks
175.0862	7.76	2.73E+02	peaks
177.0466	4.30	2.40E+02	peaks
177.0467	5.03	1.03E+03	peaks
177.0467	4.41	1.26E+02	peaks
184.0883	3.66	1.22E+03	decrease
184.0886	4.97	5.76E+01	single peak
186.0857	3.73	3.99E+02	decrease
187.0938	6.76	1.03E+02	peaks
188.1281	18.28	5.24E+02	decrease
191.0807	8.67	1.64E+02	peaks
191.0809	8.97	1.50E+02	peaks
195.0484	7.89	1.36E+02	peaks
195.8109	1.10	7.89E+02	peaks
197.8079	1.12	1.00E+03	peaks
197.8079	1.14	5.16E+02	single peak
199.0289	5.04	1.80E+02	peaks
199.8049	1.11	2.53E+02	peaks
199.8052	1.14	2.40E+02	peaks
203.0613	11.45	1.28E+02	peaks
203.0811	8.50	1.49E+02	peaks
204.1012	10.24	6.19E+02	peaks
205.0612	6.58	1.70E+02	peaks
205.0618	6.59	5.36E+02	peaks
206.1170	11.04	1.51E+02	peaks
216.0329	14.82	2.36E+02	peaks
216.0329	12.86	1.65E+02	peaks

Chapter 5 Spatial Trend detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

Table-A 5-6 continued.

218.1284	6.02	1.32E+02	peaks
219.0324	13.19	9.02E+01	peaks
220.1329	14.35	2.06E+02	peaks
223.0363	5.40	2.18E+02	peaks
224.9834	4.82	3.36E+02	peaks
232.0267	14.71	3.79E+02	peaks
232.0275	14.75	1.92E+02	peaks
232.0884	8.98	2.36E+02	peaks
234.0253	14.73	2.93E+02	peaks
234.0428	14.75	1.00E+02	peaks
234.0429	14.14	2.71E+02	peaks
234.0440	14.12	1.60E+02	peaks
234.0864	6.90	1.77E+02	peaks
234.0868	7.44	2.72E+02	peaks
235.0271	11.18	4.59E+02	decrease
236.0549	5.17	3.48E+02	peaks
236.0583	14.15	2.86E+02	peaks
236.0583	12.17	2.61E+02	peaks
237.0254	13.20	2.94E+02	peaks
237.0421	11.19	1.12E+03	decrease
237.0866	5.84	6.39E+02	decrease
239.0396	13.19	5.72E+02	peaks
239.0396	11.20	3.39E+02	peaks
239.9293	7.52	1.40E+02	peaks
241.9264	7.52	1.03E+02	peaks
243.0681	9.15	1.99E+03	peaks
243.0688	7.03	7.26E+02	decrease
243.0688	8.61	2.05E+03	peaks
244.0549	10.23	1.49E+02	peaks
246.0067	12.37	3.47E+02	peaks
246.0069	10.53	2.74E+02	peaks
246.1236	7.02	1.27E+02	peaks
248.0852	14.11	1.43E+02	peaks
249.1232	9.63	2.14E+02	peaks
251.0656	10.01	9.34E+02	decrease
251.0659	9.51	3.67E+02	decrease
253.0631	15.42	7.44E+01	peaks
255.0621	5.40	1.94E+02	single Peak
262.0294	6.20	1.05E+02	peaks
263.0288	3.31	2.25E+02	peaks
269.0237	6.06	2.03E+02	peaks
269.9039	8.54	2.35E+02	peaks
273.0482	10.01	2.06E+02	peaks
277.0443	5.40	3.51E+02	single peak
279.0928	14.70	5.98E+02	peaks
279.0929	12.96	3.49E+02	decrease
294.0013	11.09	1.12E+02	peaks
295.0920	6.40	1.58E+02	peaks
295.9896	8.07	4.27E+02	peaks
295.9901	10.06	5.55E+02	peaks
295.9902	6.03	1.64E+02	peaks
295.9904	9.33	1.89E+02	peaks
297.2771	25.51	1.21E+03	peaks
297.9874	10.07	1.81E+02	peaks
298.9421	10.80	2.42E+02	peaks
299.1313	14.84	1.90E+02	single peak
299.1314	13.35	1.01E+03	peaks
299.1321	13.07	3.96E+02	peaks
299.1322	17.88	2.72E+02	peaks
301.0753	14.70	1.50E+02	peaks
301.9323	9.17	2.94E+02	peaks
301.9324	9.47	3.02E+02	peaks
302.0710	15.39	1.79E+02	peaks
303.9299	9.47	3.29E+02	peaks
303.9299	9.16	3.99E+02	peaks
303.9304	7.32	3.22E+02	peaks



Chapter 5 Spatial Trend detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

Table-A 5-6 continued.

304.0683	15.38	1.82E+02	peaks
304.1037	4.78	1.26E+02	peaks
305.9281	9.47	9.65E+01	peaks
306.9311	12.77	3.95E+02	peaks
308.2064	4.86	1.57E+02	peaks
308.9283	12.79	1.98E+02	peaks
309.0728	7.57	4.08E+02	single peak
309.0732	6.50	3.31E+02	peaks
309.0734	5.44	1.44E+03	single peak
310.1494	2.70	4.34E+02	peaks
312.0297	7.09	2.02E+02	peaks
312.2532	17.98	1.94E+02	decrease
312.3257	20.78	4.56E+02	peaks
312.3614	22.25	1.17E+03	peaks
315.1259	12.50	4.52E+02	peaks
315.1267	10.62	5.02E+02	decrease
315.1271	12.84	4.99E+02	peaks
318.8327	1.11	3.44E+02	peaks
323.0673	12.44	5.16E+02	peaks
324.1283	1.83	2.38E+02	peaks
324.1285	2.64	2.39E+02	peaks
324.1342	13.30	2.29E+02	peaks
327.0834	6.51	8.01E+02	single peak
327.0838	5.44	2.91E+02	single Peak
330.2642	17.93	2.14E+02	decrease
338.1671	14.70	2.68E+02	peaks
339.0824	7.25	3.19E+02	decrease
339.0825	8.03	4.48E+02	decrease
341.0972	7.24	5.32E+02	decrease
341.0975	7.98	3.17E+02	decrease
341.0985	7.57	1.22E+03	single Peak
341.0987	8.95	4.62E+02	peaks
345.1080	12.43	1.18E+02	peaks
346.1110	2.64	3.11E+02	peaks
349.0658	6.51	4.00E+02	peaks
349.0659	5.44	1.52E+03	single Peak
352.2327	5.54	1.93E+02	peaks
363.0796	7.24	2.92E+02	peaks
363.0809	7.57	5.18E+02	single peak
367.0769	6.08	7.20E+02	decrease
368.1549	3.11	2.31E+02	peaks
369.0922	6.08	1.20E+03	decrease
369.0927	6.87	1.30E+02	peaks
377.1050	4.88	2.21E+02	peaks
389.3373	17.95	4.19E+02	decrease
391.0748	6.08	3.68E+02	peaks
400.1726	7.57	2.23E+02	peaks
408.3080	19.76	4.32E+02	decrease
465.0634	14.73	2.70E+02	peaks
474.9847	11.07	2.49E+02	peaks
492.4392	28.59	4.75E+02	peaks
511.5192	19.12	3.51E+02	decrease
530.0635	15.17	6.11E+02	peaks
659.5203	18.21	1.71E+03	decrease
663.4515	28.50	1.22E+03	peaks
681.5013	17.93	1.53E+03	decrease
681.9812	14.54	9.71E+02	peaks

## Chapter 5 Spatial Trend detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

Table-A 5-7 – Features which partially withstand the treatment process compared to the WWTP influent (SP 3) and their corresponding trends. In (a) the features of the WWTP influent are shown which have an increasing trend, in (b) these features of the WWTP influent which have a decreasing trend and in (c) the corresponding features of the WWTP effluent with their trends are listed describing these which withstand treatment process.

(a) SP 3: Features with an increasing trend.

Feature ID	Trend	m/z	Retention time in min
F1719	increasing	344.2275	6.97
F1808	increasing	371.2275	7.71
F1870	increasing	388.2535	7.72
F1937	increasing	415.254	8.36
F1978	increasing	432.28	8.36
F2031	increasing	459.2804	8.94
F2062	increasing	476.3063	8.94
F2131	increasing	520.3326	9.46
F2136	increasing	523.0462	5.27
F2161	increasing	537.0352	5.26
F2193	increasing	552.0001	5.27
F2200	increasing	553.0063	5.26
F2223	increasing	564.3586	9.92
F2268	increasing	608.3848	10.34

(c) SP 5: Features which partially withstand the treatment.

Feature ID	Trend	m/z	Retention time in min
F916	decrease	216.0329	6.86
F954	monotone	220.0450	12.85
F1117	peak	244.1905	19.94
F1975	decrease	743.4744	18.21

(b) SP 3: Features with a decreasing trend.

Feature ID	Trend	m/z	Retention time in min
F2390	decreasing	77.0396	3.53
F2400	decreasing	78.9587	2.09
F92	decreasing	84.0557	2.2
F153	decreasing	91.0541	1.76
F2448	decreasing	93.0344	6.08
F191	decreasing	98.9839	10.05
F192	decreasing	98.984	2.12
F229	decreasing	100.0755	3.44
F2531	decreasing	118.0296	5.55
F2532	decreasing	118.0297	4.38
F362	decreasing	121.0284	3.57
F2542	decreasing	121.0293	3.54
F398	decreasing	125.0153	3.67
F469	decreasing	129.0213	2.03
F2576	decreasing	137.0242	6.1
F570	decreasing	137.0963	17.89
F622	decreasing	147.1016	2.69
F710	decreasing	151.0284	3.56
F725	decreasing	152.0505	5.93
F2616	decreasing	153.0318	2.11
F747	decreasing	155.0468	10.04
F806	decreasing	160.0391	5.01
F837	decreasing	163.0389	5.86
F2681	decreasing	165.0189	3.53
F858	decreasing	167.0339	3.56
F894	decreasing	170.0411	4.81
F2706	decreasing	171.0115	2.72
F924	decreasing	173.0209	4.54
F935	decreasing	175.0152	8.88
F939	decreasing	175.0154	15.44
F959	decreasing	178.0496	5.58
F960	decreasing	178.0498	7.98
F2734	decreasing	179.0346	5.85

Chapter 5 Spatial Trend detection of LC-HRMS Data to Assess Processes in an Industrial Wastewater Treatment Plant

Table-A 5-7 continued.

F1006	decreasing	184.0887	3.67
F1020	decreasing	186.0858	3.67
F1032	decreasing	186.1487	17.89
F2745	decreasing	187.0509	4.33
F1041	decreasing	189.0158	3.55
F1059	decreasing	190.0473	4.8
F2755	decreasing	193.0997	12.25
F1072	decreasing	194.115	17.88
F1097	decreasing	198.1277	5.42
F1120	decreasing	200.2371	7.28
F1140	decreasing	205.0609	9.33
F2825	decreasing	209.0942	10.08
F1186	decreasing	211.1095	10.05
F2837	decreasing	212.9862	2.19
F2841	decreasing	214.0441	9.19
F2846	decreasing	216.0328	6.86
F2858	decreasing	218.0301	12.84
F1197	decreasing	218.0478	12.89
F1206	decreasing	220.0452	12.89
F2864	decreasing	221.118	11.29
F1214	decreasing	222.076	5.58
F1239	decreasing	231.078	15.46
F1240	decreasing	231.0781	8.88
F1245	decreasing	231.214	13.75
F1253	decreasing	233.0915	10.05
F1285	decreasing	237.0865	5.86
F1286	decreasing	237.0866	10.02
F1288	decreasing	237.0867	10.39
F1307	decreasing	242.2838	9.18
F1319	decreasing	244.1902	19.89
F1348	decreasing	251.0467	15.23
F1349	decreasing	251.0468	8.18
F1380	decreasing	259.0689	10.39
F1381	decreasing	259.0689	10.02
F1416	decreasing	267.1715	15.78
F1457	decreasing	281.1128	7.01
F1459	decreasing	281.1132	7.9
F1491	decreasing	287.1403	15.45
F1538	decreasing	299.1237	5.58
F1591	decreasing	307.1093	15.23
F1600	decreasing	309.1226	15.46
F1617	decreasing	313.1389	8.92
F1618	decreasing	313.139	8.34
F1619	decreasing	313.1394	7.73
F1686	decreasing	335.1212	8.34
F1687	decreasing	335.1214	8.9
F1713	decreasing	341.1337	6.55
F1714	decreasing	341.1338	7.57
F1717	decreasing	343.1497	6.34
F1729	decreasing	346.2142	15.45
F1752	decreasing	355.0425	3.54
F1778	decreasing	363.1161	6.53
F1947	decreasing	421.2118	10.07
F2109	decreasing	503.3064	8.02
F2135	decreasing	523.0456	3.55
F2159	decreasing	537.0349	3.54
F2190	decreasing	551.9988	3.55
F2198	decreasing	553.005	3.54
F2222	decreasing	564.3586	8.5
F2269	decreasing	608.3852	8.94

Table-A 5-7 continued.

F2307	decreasing	652.4113	9.33
F2359	decreasing	743.4725	18.16

## Chapter 6 Identification of Unknowns in Industrial Wastewater using Offline 2D Chromatography and Non-Target Screening

*Redrafted from Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. Sci. Total Environ. 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>*

### 6.1 Introduction

The monitoring of unknown organic contaminants in wastewater based on non-target approaches is of growing interest to ensure good water quality (Bader et al., 2017, 2016; Deeb et al., 2017; Hollender et al., 2017; Montes et al., 2017; Rapp-Wright et al., 2017; Ruppe et al., 2018; Schmidt, 2018). Industrial wastewaters contain a large diversity of pollutants arising from production plants or further transformation reactions in the wastewater in a complex matrix (Saghafi et al., 2019; Yu et al., 2019). Therefore, for operators of wastewater treatment plants (WWTPs), monitoring of the influent wastewater is essential to check whether pollutants possibly impair the biological treatment step and to ensure environmental protection. High-resolution mass spectrometry coupled to liquid chromatography (LC-HRMS) is nowadays commonly used for detecting and monitoring organic compounds in surface and wastewater (Bader et al., 2016; Montes et al., 2017). There are several HRMS detectors such as quadrupole time-of-flight (qTOF) analysers, orbitraps, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR) or sector field-MS. HRMS detectors allow obtaining an accurate mass which reduces the selection of possible predicted empirical molecular formulae of unknown compounds for further identification in the absence of reference standards and without prior information (like target or suspected-target lists) (Martínez Bueno et al., 2007). The interpretation of fragmentation patterns of HR-MS/MS can lead to successful structure identification, even for the lower ng L<sup>-1</sup> concentration levels, although discrimination between isomeric structures may fall short (Kern et al., 2009). Hetzel et al. and Salcedo et al. demonstrated, for example, the requirement of additional efforts beyond screening methods in enabling a chromatographic separation for discrimination of isomers (Hetzel et al., 2015; Salcedo et al., 2019). For identification, extracting the accurate mass of expected ions from the chromatogram, selecting features of sufficient intensity, checking the plausibility of the retention time and interpreting mass spectra has to be performed (Kern et al., 2009). Afterwards, specific information on each feature can be used to confirm plausible identities.

This, however, describes the ideal case. Compound identification of screened unknowns is still difficult, time consuming and not always successful (Milman and Zhurkovich, 2017; Schymanski et al., 2014). The interpretation of the mass spectra of unknowns is complicated by non-specific fragmentation (e.g. the loss of water) and rearrangement reactions (Levsen et al., 2000). Furthermore, the use of free available chemical databases including e.g. *ChemSpider* (Royal Society of Chemistry, 2017), *FOR-IDENT* (Grosse and Letzel, 2017), *US EPA* dashboard (U.S. Environmental Protection agency, n.d.), *NORMAN* SusDat (Aalizadeh et al., 2018), *MassBank* (Schymanski et al., 2019) and *NIST*, as well as *in silico* fragmentation tools such as *MetFrag* (Ruttkies et al., 2016) and *METLIN* (Guijas et al., 2018) for the identification of frequently observed organic contaminants are helpful, but nevertheless limited (Aceña et al., 2015). In addition, chemical, photochemical and microbiological transformation processes may lead to additional and unexpected products (Fenner et al., 2013). This results in a large diversity of site specific, suspected and unknown components, which complicates the identification using spectral libraries. Recently, prioritisation methods have been established in non-target screening (NTS), in order to reduce the amount of data generated and in order to prioritise components with higher relevance for subsequent identification (Blum et al., 2017; Hinnenkamp et al., 2019; Hohrenk et al., 2019; Park et al., 2018). Commonly used methods are for example effect-directed analysis (EDA) (Blum et al., 2017), trend analysis (Hohrenk et al., 2019) and suspected target screening (Park et al., 2018). In this study, frequently observed unknown compounds, are identified with the presented offline two-dimensional (2D) chromatographic method coupled to NTS. In contrast to common identification methods of untargeted compounds, in this study, only a fraction containing the unknown of interest is analysed. This serves as prioritisation of the large diversity of unknowns as well as purification of sample and provides a simplified basis for compound identification. Industrial wastewater samples are routinely screened for organic contaminants by LC-UV. Unknown analytes which have been observed several times during a long-term LC-UV monitoring are added to an in-house spectral database library. The '*known unknowns*' serve as prioritised compounds to demonstrate the feasibility of the approach presented here.

## 6.2 Experimental Section

### 6.2.1 Chemicals and Reagents

All internal standards (ISTD) were of analytical grade with a purity of  $\geq 95\%$ . The corresponding chemical abstract service (CAS) numbers are listed in the supporting information (see Table-A 6-1). Methanol with LCMS grade was purchased from Honeywell Riedel-de-Haën™ (Seelze,

Germany), a Milli-Q® (Q-PoD®) ultra-pure water system from Merck KGaA (Darmstadt, Germany) was used and formic acid 99% was purchased from Fisher Chemical (Geel, Belgium).

## 6.2.2 Sample Material

24-h composite flow industrial wastewater samples (directly from the drain of the industrial plant, the WWTP influent, the WWTP effluent and samples within the WWTP) were collected for analysis. All samples were collected in pre-cleaned 250 mL high-density polyethylene (HDPE) bottles and stored at 8 °C until analysis.

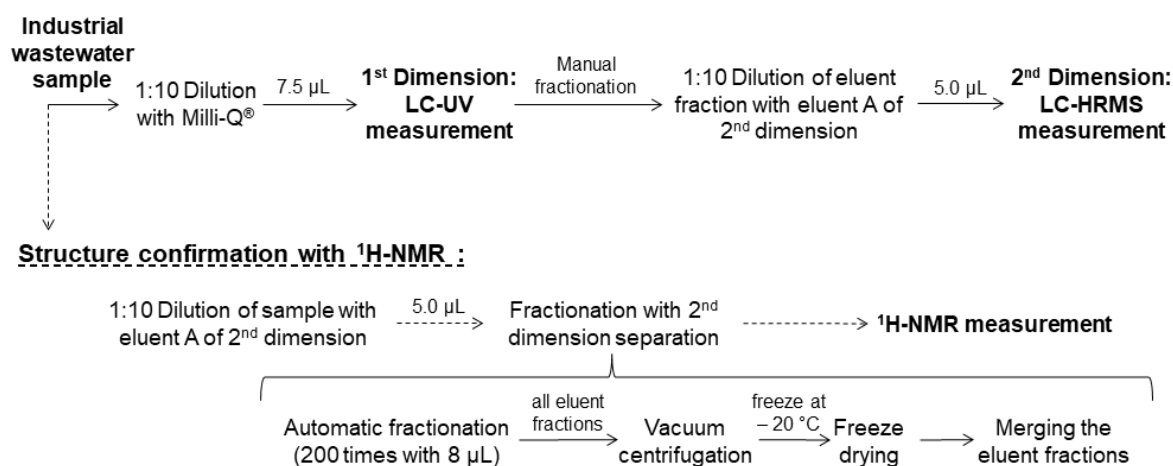


Figure 6-1 – Flowchart of the industrial wastewater sample analysed by the offline 2D LC system. For structure confirmation, the wastewater sample was measured again for enrichment using LC-HRMS and analysed by <sup>1</sup>H-NMR.

In both separation dimensions quality controls (see QC in Table-A 6-1) were measured to check the separation efficiency, the column quality and as positive control to validate the data processing workflow.

## 6.2.3 Analytical Methods

### 6.2.3.1 LC Separation and UV Detection in the First Dimension

Unfiltered operation plant samples were diluted by a factor of 10 with Milli-Q® water before analysis. All LC-UV data were collected using an Agilent 1260 Infinity II (Waldbronn, Germany). The separation was carried out on a Purospher® STAR RP end-capped column (C18 column, 125 mm x 4.0 mm, 3.0 µm particles, Merck, Darmstadt, Germany) with the corresponding pre-column. The injection volume was 7.5 µL and the flow rate 750 µL min<sup>-1</sup>. The mobile phase

consisted of salt solution (5 mM potassium dihydrogen phosphate and 25 mM sodium sulphate, pH = 4.5, A) and acetonitrile (B). The following solvent gradient was applied for eluent B: in 16 min rise from 0% to 8.0%, in the next 25 min rise to 22%, followed by a further increase to 74% within 13 min, hold for 7 min, in 1 min decrease to 0.0% and re-equilibration for 6 min. The DAD spectra were recorded from 210 to 598 nm. Peaks of interest were manually fractionated after this dimension (see Figure 6-1).

The '*known unknown*' of interest of the original sample was isolated offline by collecting the eluent fraction after UV detection on the first dimension in LC-vials (see Figure 6-1) for sample purification to simplify compound identification. Initially, the LC-UV system was equilibrated to provide stable retention time of the compounds. After three injections with comparable retention time and a standard deviation of 0.05 min, the eluate of the LC was collected after the UV detection and before the waste (see Figure 6-1). The signal was cut at full width at half maximum (FWHM, Figure is presented in Figure-A 6-5), which resulted in an average eluent fraction volume of 200  $\mu$ L. The cutting process was optimised during validation (For validation, wastewater samples were spiked with target analytes of Table-A 6-1). In addition, a blank was fractionated for evaluation to correct background signals.

#### 6.2.3.2 LC-HRMS Separation and Detection in the Second Dimension

The elution fraction collected in the first dimension (see Figure 6-1) was subsequently analysed by LC-HRMS after dilution by a factor of 10 with mobile phase A of the second dimension. For HPLC, an Agilent 1290 Infinity HPLC system (Waldbronn, Germany) equipped with a Raptor™ Ultra Aqueous C18 analytical column (100 mm x 2.1 mm, 3.0  $\mu$ m particle size, Restek GmbH, Bad Homburg v. d. Höhe, Germany) and Restek™ Trident cartridge (10 x 2.1 mm) and filter (2 mm, 0.5  $\mu$ m) was used. The (pre-)column was changed depending on the peak shape of the analytes of the quality control (QC, see SI 1 b). Gradient elution was performed with 0.1% formic acid in Milli-Q® water (A) and 0.1% formic acid in methanol (B). The initial mobile phase composition (0% B) was held constant for 0.50 min, reached 95% B in 19 min, held there for 6.0 min and finally re-equilibrated for 6 min (0% B). Flow rate and the injection mode were set to 300  $\mu$ L min<sup>-1</sup> and 5  $\mu$ L. For HRMS, a x500R qTOF with electrospray ion source Turbo V™ (SCIEX GmbH, Darmstadt, Germany) was used. Positive as well as negative electrospray ionisation was applied. The resolution of the instrument is shown in the supporting information (see Figure-A 6-1). Full scan HRMS data were recorded within a mass range of mass-to-charge ratio ( $m/z$ ) 70 – 950 for each sample. For MS/MS, precursor ions exceeding an intensity of 50 cps were used for data-dependent acquisition (DDA). Up to 10 data-dependent MS/MS scans in each cycle (0.56 s) covering a mass range of  $m/z$  30 – 700, were obtained. The full-scan data were processed with SCIEX OS Software '*Non-targeted*

*screening*' workflow (Version 1.4.1). For formulae identification, the Formula Finder of SCIEX OS Software was used. For MS/MS interpretation, the *in silico* fragmentation tool *MetFrag* was used, which is a freely available software for the annotation of high precision tandem mass spectra (Ruttkies et al., 2016).

### 6.2.3.3 <sup>1</sup>H-NMR Analysis for Structure Confirmation

For the analysis via <sup>1</sup>H-NMR spectroscopy, the unknown of interest was fractionated of the original wastewater sample using LC-HRMS (retention time window of 8 s) (see Figure 6-1). For the identification by NMR, a sufficient amount of the unknown compound (which is significantly larger than the volume of one single fraction) is necessary. Therefore, multiple separations of the same sample and collection of the same fraction were applied. For the fraction collection, the Fraction Collector III (Waters GmbH, Eschborn, Germany) was used. The diluted sample (1:5, V/V with Milli-Q®) were injected 200 times with 8 µL injection volume to increase the amount of unknown compound. Thereafter, the extracts were concentrated by evaporating the volatile solvents in a SpeedVac (Eppendorf Concentrator plus, eppendorf, Hamburg, Germany) and freezing at -20 °C. Subsequently, the extracts were freeze-dried with high vacuum (Christ Alpha 2-4 LSCbasic, Osterode am Harz, Germany). The freeze-dried elution fractions, as well as reference material (3,4-dichloro-2,6-dinitrophenol), were dissolved in deuterated methanol and transferred to a 3 mm NMR tube for analysis. Samples were analysed using a Bruker AV III 700 NMR equipped with 5 mm TCI cryo probe (Bruker, Massachusetts, USA) and larmor frequency of 700 MHz.

## 6.3 Results and Discussion

### 6.3.1 Fractionation of 'Known Unknown' in the First Dimension

Standard monitoring of the wastewater is applied by LC-UV and regularly observed signals are added to a database including all '*known unknowns*'. One of the '*known unknowns*' was selected to demonstrate the prioritisation workflow. The compound elutes after 41.1 min and has its maximal UV absorption at 233 nm (see Figure 6-4 a). It was fractionated in the first-dimension eluent after three sample's injection with comparable retention times and a standard deviation of 0.05 min.

### 6.3.2 Identification Process of prioritised Compound with LC-HRMS Analysis of the Selected LC-UV Fraction

By analysing the eluent fraction of the first dimension with LC-HRMS, 1610 features were detected in negative ionisation mode (see Table 6-1) and 3170 features were detected in



positive ionisation mode. So, despite fractionation, there were many features detected, which obviously did not belong to the peak of interest. Of those, many features were present in the blank as well and therefore could be omitted. Checking for false positives (e.g. peak shape, invalid integration of noise) and adducts resulted in a further reduction of possible features (see Table 6-1).

Table 6-1 – Data evaluation workflow of detected features filtration for positive (ESI (+)) and negative ionisation (ESI (-)) mode; RT  $\triangleq$  retention time, auto  $\triangleq$  software-based filter, manual  $\triangleq$  manual exclusion of features.

<b>Filter criteria parameters</b>	<b>Detected features of Elution fraction</b>	
	<b>ESI (+)</b>	<b>ESI (-)</b>
Initial ( <i>auto</i> ):		
• RT window: 1 – 26 min	3170	1610
• Peak detection sensitivity: exhaustive*		
Blank subtraction ( <i>auto</i> ):		
• Intensity threshold for unknown/blank: 1000	1585	805
Substitution of features without MS <sup>2</sup> data ( <i>auto</i> ):	66	13
Elimination of false positives due to peak shape criteria ( <i>manual</i> ):	1	6
Adduct grouping ( <i>auto</i> ):	1	5

\*Manufacturer recommended setting.

In negative ionisation mode, five relevant features remained (see Table 6-2). The feature with ID 4 showed the highest intensity. The other features having the same retention time of 11.17 min proved to belong to this feature as isotopes (ID 2 – ID 5, same peak shape,  $m/z$  250.9262  $\triangleq$  [M-H]<sup>-</sup>,  $m/z$  251.9310  $\triangleq$  <sup>13</sup>C of [M-H]<sup>-</sup>,  $m/z$  252.9239  $\triangleq$  [M-H]<sup>-</sup> with <sup>37</sup>Cl isotope,  $m/z$  234.9326  $\triangleq$  loss of an oxygen by insource reduction), which reduced the number of features of interest to two. In positive ionisation mode, after exclusion of false positive hits and adduct grouping, only one feature was left ( $m/z$  325.0832, RT = 10.62 min and Intensity = 2.00E+04 cps). Because of retention time, accurate mass and significant fragments, this feature was assigned to belong to the same compound as the feature with ID 1 in negative ionisation mode.

Table 6-2 – Remaining features in negative ionisation mode (ESI (-)) of non-target LC-HRMS analysis after applying blank subtraction, exclusion of features without MS<sup>2</sup> data and false positives, adduct grouping and retention time filtering.

ID	Precursor Mass in Da	Retention Time in min	Height in cps
1	323.0690	10.63	2.78E+03
2	234.9326	11.17	2.06E+03
3	252.9239	11.17	3.89E+04
4	250.9265	11.17	6.02E+04
5	251.9310	11.17	6.55E+03

Using the MS software, empirical formulae were evaluated for the two remaining features. SCIEX OS Formula Finder was configured for 'Man-Made Compounds' with maximal elements of C<sub>0-49</sub>, H<sub>0-75</sub>, Cl<sub>0-5</sub>, F<sub>0-3</sub>, N<sub>0-10</sub>, O<sub>0-16</sub>, P<sub>0-1</sub>, S<sub>0-3</sub> and a mass tolerance of 10 ppm (manufacturer recommended setting). During analysis, the settings were adapted, taking the isotope pattern and information of the operation plant of relevant chemical element compositions into account (C<sub>0-49</sub>, H<sub>0-75</sub>, O<sub>0-5</sub>, N<sub>0-5</sub>, Cl<sub>0-2</sub>, S<sub>0-3</sub>, with a mass tolerance of 5 ppm). This results in no empirical molecular formulae for the feature with ID 1 (see Table 6-2). Consequently, this feature was excluded, too.

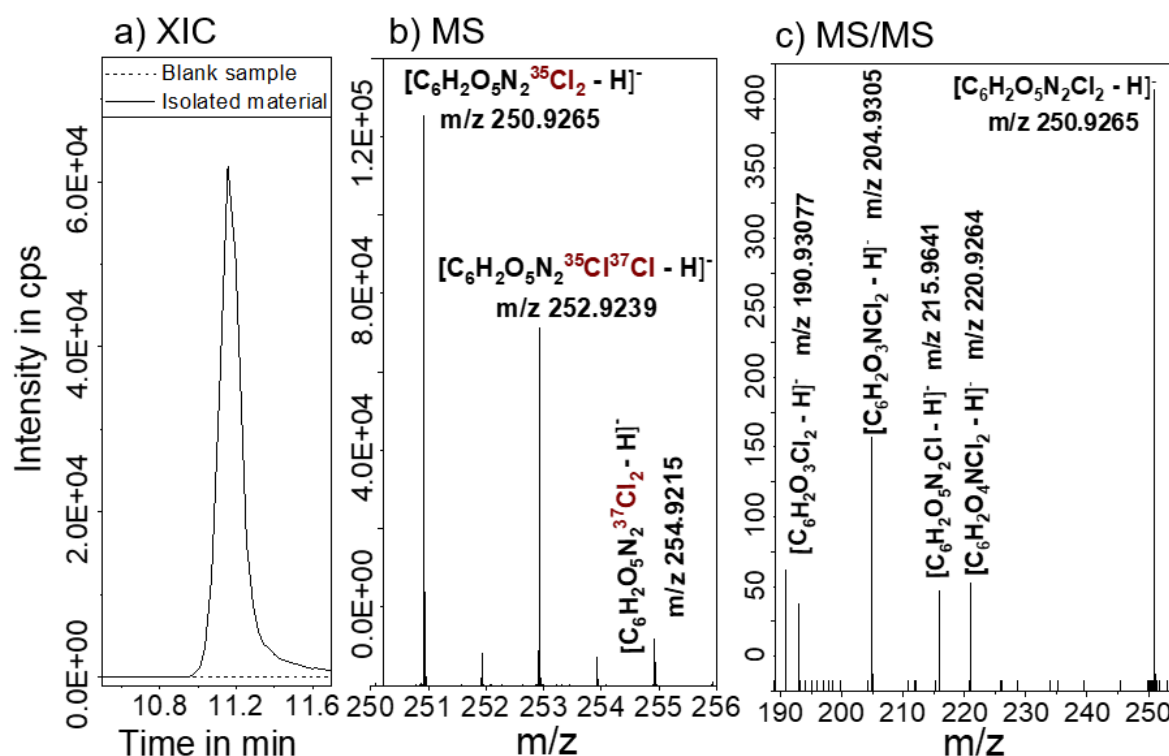


Figure 6-2 – LC-HRMS results in ESI (-) of the feature with ID 4; (a) Signal intensity comparison of feature 4 (XIC m/z 250.9265) for eluent fraction of wastewater and blank sample; (b) MS spectrum of precursor-ion of m/z 250.9265 with isotopic pattern for di-chlorination and (c) MS/MS spectrum of precursor-ion of m/z 250.9265 of eluent fraction.

The prioritisation of offline 2D LC should lead to the highest intensity of the prioritised compound, the 'known unknown' in LC-HRMS. Therefore, the further identification focus was

on the feature with ID 4 with the highest intensity (see Figure 6-2 a). Figure 6-2 b shows a part of the mass spectrum which includes the isotopic pattern of di-chlorination ( $^{35}\text{Cl}_2$ : 100%,  $^{35}\text{Cl}^{37}\text{Cl}$ : 63.99%,  $^{37}\text{Cl}_2$ : 10.24%). Figure 6-2 c shows the MS/MS spectrum of the unknown substance with fragments confirming the di-chlorination. Both MS and MS/MS are definite signs of the later suggested dichlorodinitrophenol. The empirical formulae of Table 6-3 were predicted for the feature with ID 4.

Table 6-3 – Excerpt of empirical formulae of the feature with ID 4 assigned by Formula Finder (SCIEX OS) in consideration of operation plant information with maximal element options of C<sub>0-30</sub>, H<sub>0-30</sub>, O<sub>0-5</sub>, N<sub>0-5</sub>, Cl<sub>0-3</sub>, S<sub>0-3</sub> and a mass tolerance of 5 ppm.

Formula	Score	m/z	Error in ppm	Error MS/MS in ppm
C <sub>6</sub> H <sub>2</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>5</sub>	81.7	250.9262	1.1	3.2
C <sub>7</sub> H <sub>6</sub> Cl <sub>2</sub> N <sub>2</sub> S <sub>2</sub>	59.3	250.9265	4.6	3.0

Testing the proposed formulae with *MetFrag*, for C<sub>7</sub>H<sub>6</sub>Cl<sub>2</sub>N<sub>2</sub>S<sub>2</sub> two possible compounds 4,5-dichloro-1H-imidazole-2-carbodithioate and 5,7-dichloro-4-ethyl-thienothiadiazine were suggested. The corresponding fragments were not in accordance with the MS/MS spectrum of Figure 6-2 c, confirming only the chlorine fragment with m/z 34.9689. For the proposal of C<sub>6</sub>H<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>5</sub>, *MetFrag* processed three candidates (see Figure 6-3); 3,6-dichloro-2,4-dinitro-phenol, 3,4-dichloro-2,6-dinitro-phenol and 2,3-dichloro-4,6-dinitro-phenol.

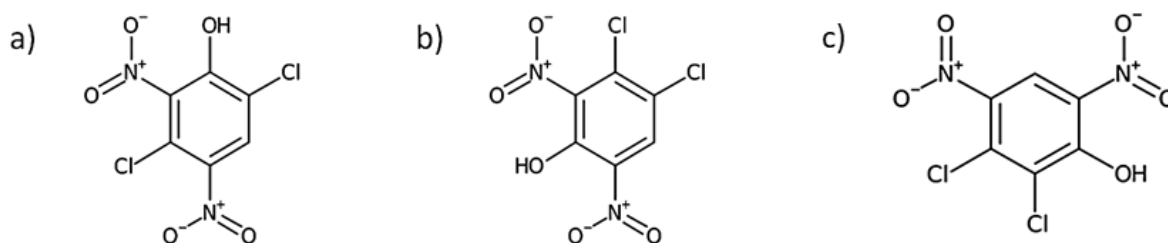


Figure 6-3 – Proposed structures for formula finder result C<sub>6</sub>H<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>5</sub> for the feature with ID 4: (a) 2,5-dichloro-4,6-dinitro-phenol, (b) 3,4-dichloro-2,6-dinitro-phenol and (c) 2,3-dichloro-4,6-dinitro-phenol.

The suggested dichlorodinitro substituted phenols for C<sub>6</sub>H<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>5</sub> were suitable candidates. The isotope pattern of the di-chlorination (see Table 6-2 b), as well as the MS/MS spectra (see Table 6-2 c), the calculated double bond equivalent (DBE) of six (five double bonds and one ring structure, see 6.6.2.1) and the ionisation in the negative mode support this assumption (see Figure 6-2). To verify the hypothesis, the unknown's retention times and spectra (UV, MS, MS/MS) were compared with the purchased reference substance 3,4-dichloro-2,6-dinitrophenol (only commercially available isomer). The retention times of the reference material in both, the first dimension (41.1 min) and the second dimension (11.2 min), confirmed the retention times of the unknown compound.

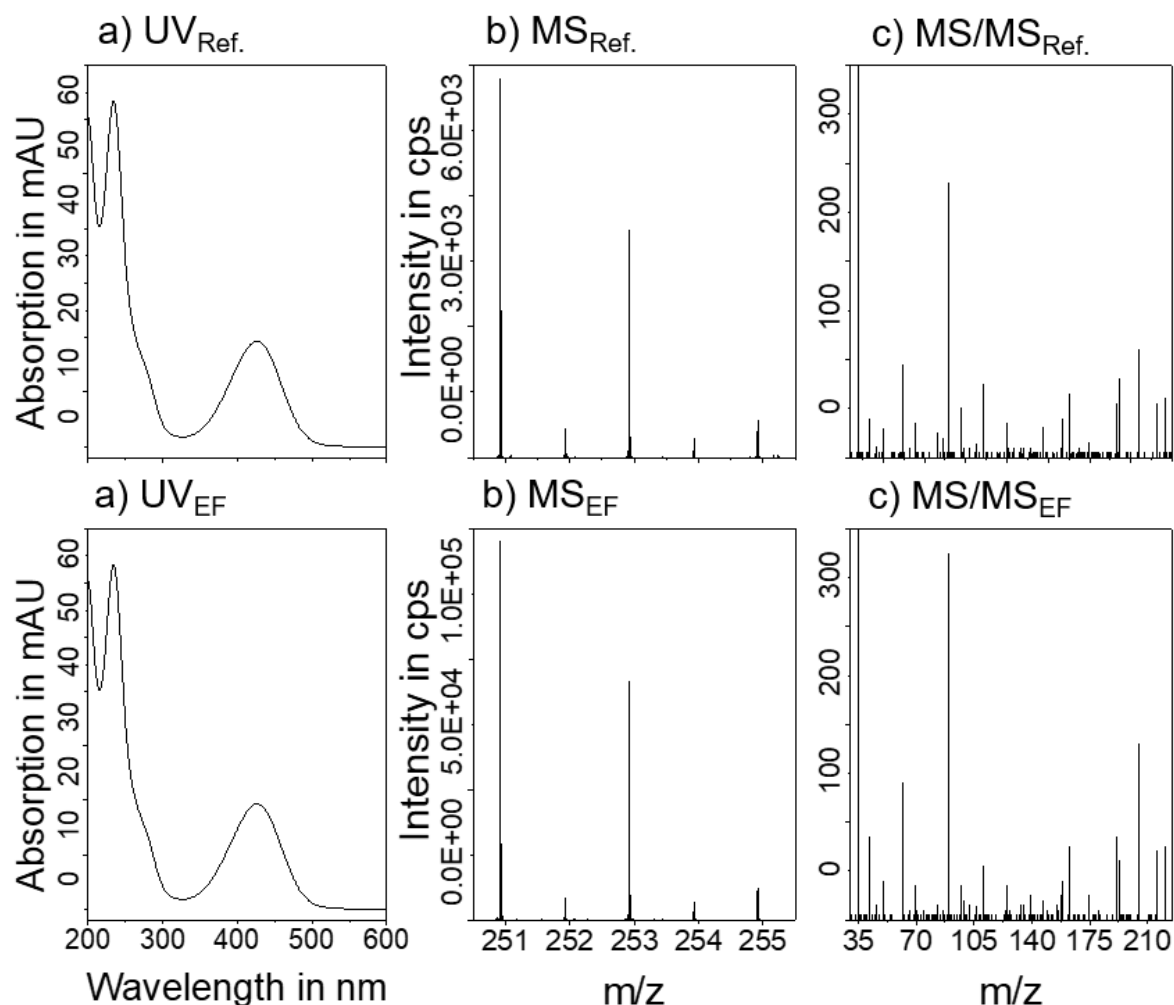


Figure 6-4 – Comparison of authentic standard 3,4-dichloro-2,6-dinitrophenol (upper row,  $50 \mu\text{g L}^{-1}$ ) and eluent fraction of wastewater sample (lower row); UV spectra of (a) authentic standard (Ref.) and (d) of eluent fraction (EF); cut-out of MS spectrum of (b) precursor-ion  $m/z$  250.9265 of authentic standard and (e) of precursor-ion  $m/z$  250.9265 of eluent fraction showing identical isotopic pattern; MS/MS spectrum of (c) precursor-ion  $m/z$  250.9265 of authentic standard and (f) of precursor-ion  $m/z$  250.9265 of eluent fraction.

Additionally, the UV spectrum of the reference material was in agreement with the unknown compound (see Figure 6-4 a & d). The match factor of the UV spectra was 999.655 (ChemStation). Furthermore, the MS data of the unknown and the authentic standard perfectly matched. Both MS spectra showed the isotopic pattern of di-chlorination. The MS/MS spectrum of the unknown compound agreed with the authentic standard by 99.2% (SCIEX OS). The highest fragments with  $m/z$  220.9264 ( $[\text{C}_6\text{H}_2\text{O}_4\text{NCl}_2 - \text{H}]^-$ ),  $m/z$  215.9641 ( $[\text{C}_6\text{H}_2\text{O}_3\text{N}_2\text{Cl} - \text{H}]^-$ ),  $m/z$  204.9305 ( $[\text{C}_6\text{H}_2\text{O}_3\text{NCl}_2 - \text{H}]^-$ ) and  $m/z$  190.93077 ( $[\text{C}_6\text{H}_2\text{O}_3\text{Cl}_2 - \text{H}]^-$ ) are present in both MS/MS spectra. The mass spectra of reference and 'known unknown' match each other. The difference in intensity of mass spectra is explained due to different measured concentrations ( $50 \mu\text{g/L}$  for authentic standard 3,4-dichloro-2,6-dinitrophenol). In summary, all MS-, MS/MS- and UV-spectra, as well as retention times, were consistent with the level one classification introduced by Schymanski et al. (Schymanski et al., 2014).

### 6.3.3 Identity Verification of 'Known Unknown' by $^1\text{H-NMR}$

Although, a level one classification was achieved (Schymanski et al., 2014), the exact substitution of the dichlorodinitrophenol isomer found had not been clarified. To determine the unique structure of the isomer, an additional  $^1\text{H-NMR}$  measurement was performed. Since the unknown of interest was detected and identified as dichlorodinitrophenol, the prioritisation of offline 2D LC was no longer necessary. In Figure-A 6-6 the possible constitutional isomers of dichlorodinitrophenol are listed. According to the presented  $^1\text{H-NMR}$  spectra (see Figure 6-5, extracting the blank value) the detected analytes had to be constitutional isomers of dichlorodinitrophenol (see Figure-A 6-6) but were not identical to the reference material (see Figure 6-5 a). The chemical shift of the aromatic proton in the sample (8 ppm) differed from the authentic reference material (3,4-dichloro-2,6-dinitrophenol). Furthermore, an identification of the specific isomer was not possible because of the high symmetry of dichlorodinitrophenol and the presence of only one meaningful hydrogen atom. The other signals occurring in the sample (4.5 to 0.5 ppm) did not belong to the identified 'known unknown' but correspond to the blank sample (see Figure 6-5 c). The result was in no contradiction to the level one identification, according to Schymanski et al. (Schymanski et al., 2014), because the  $^1\text{H-NMR}$  analysis is not required.

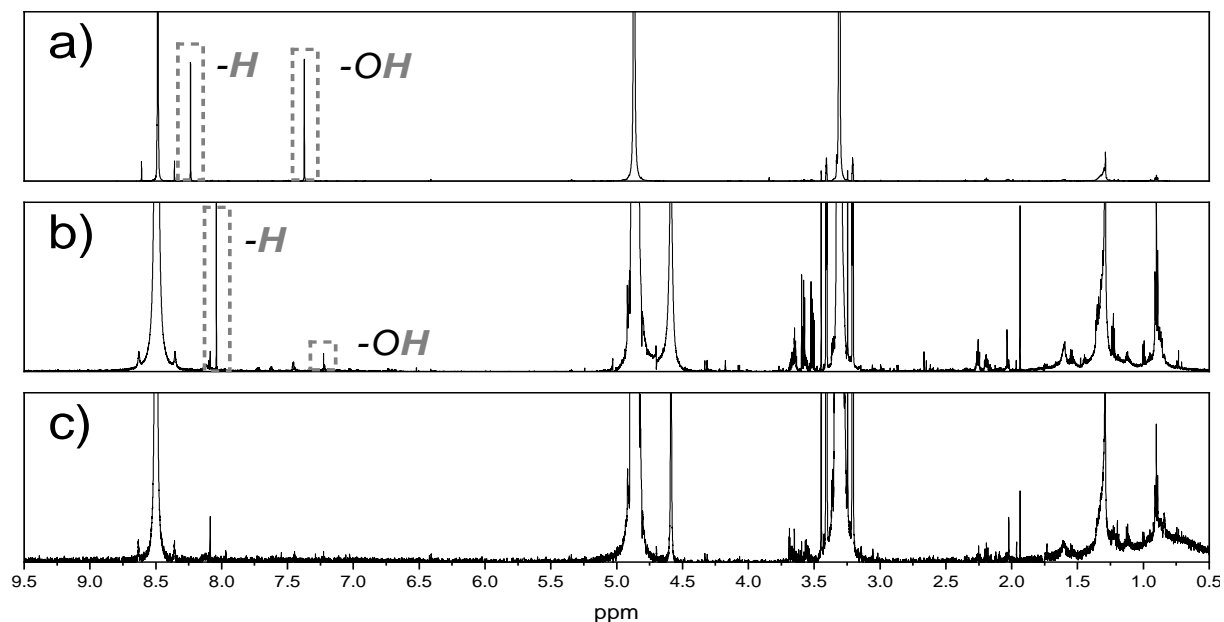


Figure 6-5 – Comparison of  $^1\text{H-NMR}$  spectra of (a) authentic standard 3,4-dichloro-2,6-dinitrophenol, (b) eluent fraction of wastewater sample and of (c) blank. The chemical shift of around 8 ppm belongs to those of aromatic compounds.

### 6.3.4 Quantitative Screening of identified '*Known Unknown*'

The unfiltered 24-h composite flow samples were analysed by LC-HRMS diluted by a factor of 10 with mobile phase A of the second dimension. The '*known unknown*' was approximately quantified retrospectively in real samples (without fractionation) using reference standard (3,4-dichloro-2,6-dinitrophenol, calibration curves see Figure-A 6-10, LOD = 100 ng L<sup>-1</sup>, LOQ = 500 ng L<sup>-1</sup>) and deuterated ISTDs (see Table-A 6-1, Mecoprop-D3 10 µg/L was used for quantification). The quantitative determination of the dichlorodinitrophenol is only an approximation because of the missing correct isomer. However, the available reference standard provides a good estimation because of its structural similarity. Furthermore, by analysing water samples of industrial plants, the WWTP and receiving water, it was shown that the unknown analyte was removed (< LOQ of 500 ng L<sup>-1</sup>) in the WWTP. The highest average concentration was observed for industrial wastewater samples with approximately 0.4 mg L<sup>-1</sup> (see Figure-A 6-11). In addition, for assessment of substance toxicity, high-performance thin-layer chromatography (HPTLC) coupled to *Aliivibrio fischeri* bioassay was performed. Details are mentioned in 6.6.1.3 and 6.6.2.3. The results confirmed some of 1960's patents, which emphasise that some chloronitrophenol pesticides may be toxic (Pyne, 1963, 1962).

## 6.4 Conclusions

A prioritisation workflow for the identification of industrial wastewater contaminants was developed to manage the large amount of data generated in an NTS. As proof of concept, the identification process of one prioritised compound was demonstrated. Although the results of LC-UV/LC-HRMS for the unknown compound confirmed the level one classification introduced by Schymanski et al. (Schymanski et al., 2014), the additional <sup>1</sup>H-NMR analysis showed that the unknown is more likely another isomer of dichlorodinitrophenol compared to the reference standard. However, the presented analytical method is suitable for identification of '*known unknowns*' as prioritised compounds from routine analysis occurring in industrial wastewater with the limitation of discrimination of some isomeric structures. Moreover, identification with retrospective quantification (based on an authentic standard) is possible. The development of new prioritisation methods capable to prioritise and identify unknown compounds in water samples is important as NTS becomes wider-more spread. Smart prioritisation strategies combining the power of LC-HRMS with further information, as presented in this approach, or advanced statistics can lead to a much better understanding of the wastewater samples and treatment processes. The analytical procedure of NTS is no longer the biggest challenge these days. However, the processing and interpretation of the data generated by the screening procedures still are. By prioritisation, data complexity can be reduced in a meaningful way.

## 6.5 References

- Aalizadeh, R., Alygizakis, N., Schymanski, E., Slobodnik, J., 2018. S0 | SUSDAT | Merged NORMAN Suspect List: SusDat. <https://doi.org/10.5281/zenodo.2664077>
- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography–high-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 407, 6289–6299. <https://doi.org/10.1007/s00216-015-8852-6>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2017. LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes. *Anal. Chem.* 89, 13219–13226. <https://doi.org/10.1021/acs.analchem.7b03037>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2016. General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Anal. Chim. Acta* 935, 173–186. <https://doi.org/10.1016/j.aca.2016.06.030>
- Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Sci. Total Environ.* 575, 265–275. <https://doi.org/10.1016/j.scitotenv.2016.09.135>
- Deeb, A.A., Stephan, S., Schmitz, O.J., Schmidt, T.C., 2017. Suspect screening of micropollutants and their transformation products in advanced wastewater treatment. *Sci. Total Environ.* 601–602, 1247–1253. <https://doi.org/10.1016/j.scitotenv.2017.05.271>
- Fenner, K., Canonica, S., Wackett, L.P., Elsner, M., 2013. Evaluating pesticide degradation in the environment: Blind spots and emerging opportunities. *Science* (80- .). 341, 752–758. <https://doi.org/10.1126/science.1236281>
- Grosse, S., Letzel, T., 2017. User Manual for FOR-IDENT Database 3, 1–35.
- Guijas, C., Montenegro-Burke, J.R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., Koellensperger, G., Huan, T., Uritboonthai, W., Aisporna, A.E., Wolan, D.W., Spilker, M.E., Benton, H.P., Siuzdak, G., 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* 90, 3156–3164. <https://doi.org/doi:10.1021/acs.analchem.7b04424>.
- Hetzel, T., Teutenberg, T., Schmidt, T.C., 2015. Selectivity screening and subsequent data evaluation strategies in liquid chromatography: the example of 12 antineoplastic drugs. *Anal. Bioanal. Chem.* 407, 8475–8485. <https://doi.org/10.1007/s00216-015-8994-6>
- Hinnenkamp, V., Balsaa, P., Schmidt, T.C., 2019. Quantitative screening and prioritization based on UPLC-IM-Q-TOF-MS as an alternative water sample monitoring strategy. *Anal. Bioanal. Chem.* <https://doi.org/https://doi.org/10.1007/s00216-019-01994-w>
- Hohrenk, L.L., Vosough, M., Schmidt, T.C., 2019. Implementation of chemometric tools to improve data mining and prioritization in LC-high resolution mass spectrometry for nontarget screening of organic micropollutants in complex water matrices. *Anal. Chem.* 2. <https://doi.org/10.1021/acs.analchem.9b01984>
- Hollender, J., Schymanski, E.L., Singer, H., Ferguson, P.L., 2017. Non-target screening with high resolution mass spectrometry in the environment: Ready to go? *Environ. Sci. Technol.* 51, 11505–11512. <https://doi.org/10.1021/acs.est.7b02184>

- Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., Hollender, J., 2009. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 43, 7039–7046. <https://doi.org/10.1021/es901979h>
- Levsen, K., Preiss, A., Godejohann, M., 2000. Application of high-performance liquid chromatography coupled to nuclear magnetic resonance and high-performance liquid chromatography coupled to mass spectrometry to complex environmental samples. *Trends Anal. Chem.* 19, 27–48. [https://doi.org/10.1016/S0165-9936\(99\)00178-8](https://doi.org/10.1016/S0165-9936(99)00178-8)
- Martínez Bueno, M.J., Agüera, A., Gómez, M.J., Hernando, M.D., García-Reyes, J.F., Fernández-Alba, A.R., 2007. Application of liquid chromatography/quadrupole-linear ion trap mass spectrometry and time-of-flight mass spectrometry to the determination of pharmaceuticals and related contaminants in wastewater. *Anal. Chem.* 79, 9372–9384. <https://doi.org/10.1021/ac0715672>
- Milman, B.L., Zhurkovich, I.K., 2017. The chemical space for non-target analysis. *Trends Anal. Chem.* 97, 179–187. <https://doi.org/10.1016/j.trac.2017.09.013>
- Montes, R., Aguirre, J., Vidal, X., Rodil, R., Cela, R., Quintana, J.B., 2017. Screening for Polar Chemicals in Water by Trifunctional Mixed-Mode Liquid Chromatography-High Resolution Mass Spectrometry. *Environ. Sci. Technol.* 51, 6250–6259. <https://doi.org/10.1021/acs.est.6b05135>
- Park, N., Choi, Y., Kim, D., Kim, K., Jeon, J., 2018. Prioritization of highly exposable pharmaceuticals via a suspect/non-target screening approach: A case study for Yeongsan River, Korea. *Sci. Total Environ.* 639, 570–579. <https://doi.org/10.1016/j.scitotenv.2018.05.081>
- Pyne, W.J., 1963. Phenolic Pesticide. United States Patent Office 3,108,927, 1–4.
- Pyne, W.J., 1962. Penolic Lamprey Larvicides. United States Patent Office 3,052,601, 1–6.
- Rapp-Wright, H., McEneff, G., Murphy, B., Gamble, S., Morgan, R., Beardah, M., Barron, L., 2017. Suspect screening and quantification of trace organic explosives in wastewater using solid phase extraction and liquid chromatography-high resolution accurate mass spectrometry. *J. Hazard. Mater.* 329, 11–21. <https://doi.org/10.1016/j.jhazmat.2017.01.008>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Ruppe, S., Griesshaber, D.S., Langlois, I., Singer, H.P., Mazacek, J., 2018. Detective Work on the Rhine River in Basel – Finding Pollutants and Polluters. *Chim. Int. J. Chem.* 72, 547. <https://doi.org/10.2533/chimia.2018.547>
- Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 1–16. <https://doi.org/10.1186/s13321-016-0115-9>
- Saghafi, S., Ebrahimi, A., Mehrdadi, N., Bidhendy, G.N., 2019. Evaluation of aerobic/anaerobic industrial wastewater treatment processes: The application of multi-criteria decision analysis. *Environ. Prog. Sustain. Energy* 1–7. <https://doi.org/10.1002/ep.13166>
- Salcedo, G.M., Kupski, L., Degang, L., Marube, L.C., Caldas, S.S., Primel, E.G., 2019. Determination of fifteen phenols in wastewater from petroleum refinery samples using a dispersive liquid–liquid microextraction and liquid chromatography with a photodiode array detector. *Microchem. J.* 146, 722–728. <https://doi.org/10.1016/J.MICROC.2019.01.075>



Schmidt, T.C., 2018. Recent trends in water analysis triggering future monitoring of organic micropollutants. *Anal. Bioanal. Chem.* 410, 3933–3941. <https://doi.org/10.1007/s00216-018-1015-9>

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. <https://doi.org/10.1021/es5002105>

Schymanski, E.L., Schulz, T., Alygizakis, N., Meier, R., 2019. S1 | MASSBANK | NORMAN Compounds in MassBank. <https://doi.org/10.5281/zenodo.2621390>

U.S. Environmental Protection agency, n.d. US EPA dashboard [WWW Document]. United States Environ. Prot. URL <https://comptox.epa.gov/dashboard> (accessed 8.27.19).

Yu, Y., Wu, B., Jiang, L., Zhang, X.X., Ren, H.Q., Li, M., 2019. Comparative analysis of toxicity reduction of wastewater in twelve industrial park wastewater treatment plants based on battery of toxicity assays. *Sci. Rep.* 9, 1–10. <https://doi.org/10.1038/s41598-019-40154-z>

## 6.6 Chapter Appendix

### 6.6.1 Supplement for Chapter Materials and Methods

#### 6.6.1.1 Instrument Resolution

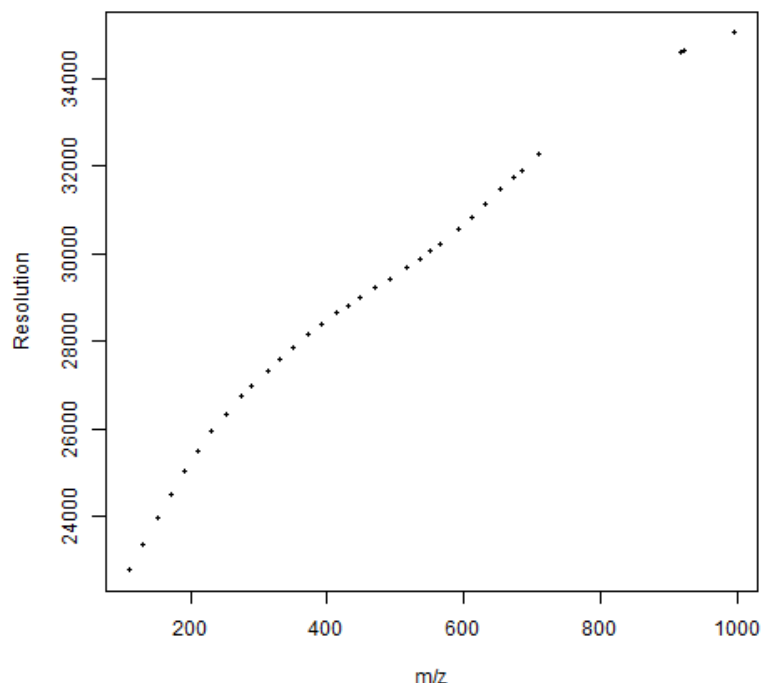


Figure-A 6-1 – Instrument Resolution of SCIEX qTOF x500R (Loos, 2018).

6.6.1.2 Used Analytes

Table-A 6-1 – Table of analytes and ISTD used for method development. Additionally, the data of quality control (QC) is listed. All data are rounded data and refer to the LC-HRMS measurement (2<sup>nd</sup> dimension). For the identification, a maximum mass error of ± 2 ppm and a maximum retention time error of ± 0.05 min were approved. Fragment ions are listed in decreasing intensity from 1 to 5. RT ≙ Retention time, Rec. ≙ Recovery (\* industrial chemical).

Target compound	Main Ion / Adduct	Sum formula	Molecular Mass in g*mol <sup>-1</sup>	Expected Mass in Da	Accurate Mass in Da	Mass Error in mDa	Mass Error in ppm	RT in min	Rec. in %	Fragment ions in Da					CAS Number
										1	2	3	4	5	
Benzamide oxime	[M+H] <sup>+</sup>	C <sub>7</sub> H <sub>8</sub> N <sub>2</sub> O	136.064	137.071	137.0709	-0.088	-0.6	2.66	-	77.0389	104.05	51.0233	137.071	119.061	613-92-3
DEF 100*	[M+H] <sup>+</sup>	C <sub>8</sub> H <sub>7</sub> NO <sub>3</sub> S	229.043	229.043	229.0429	-0.094	-0.4	17.38	-	110.99	82.9851	-	-	-	-
Phthalic Acid	[M-H] <sup>-</sup>	C <sub>8</sub> H <sub>6</sub> O <sub>4</sub>	166.027	165.019	165.0193	-0.044	-0.3	5.2	-	77.0394	121.03	96.9602	164.898	147.8929	88-99-3
Phthalimide	[M+H] <sup>+</sup>	C <sub>8</sub> H <sub>6</sub> NO <sub>2</sub>	147.032	148.039	148.0392	-0.145	-1.0	7.15	-	130.034	102.034	75.0235	120.045	51.0239	85-41-6
2-chloro-4,6-dinitrophenol	-	C <sub>6</sub> H <sub>3</sub> ClN <sub>2</sub> O <sub>5</sub>	217.973	-	-	-	-	-	-	-	-	-	-	-	946-31-6
3,4-dichloro-2,6-dinitrophenol	[M-H] <sup>-</sup>	C <sub>6</sub> H <sub>2</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>5</sub>	251.934	250.927	250.9265	-0.333	-1.3	11.17	-	220.926	215.964	204.9305	190.931	-	1664-10-4
4-chloro-2,6-dinitrophenol	-	C <sub>6</sub> H <sub>3</sub> ClN <sub>2</sub> O <sub>5</sub>	217.973	-	-	-	-	-	-	-	-	-	-	-	88-87-9
Azoxystrobin-D4 (ISTD)	[M+H] <sup>+</sup>	C <sub>22</sub> H <sub>13</sub> D <sub>4</sub> N <sub>3</sub> O <sub>5</sub>	407.142	408.149	408.1487	-0.467	-1.1	14.47	-	377.128	349.134	-	-	-	-
Bezafibrate-D6 (ISTD)	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	C <sub>19</sub> H <sub>14</sub> D <sub>6</sub> ClNO <sub>4</sub>	367.146	368.153 / 366.139	368.1527 / 366.1380	0.657 / 0.471	0.7	14.09	-	138.9953 / 274.0642	121.0655 / 154.0067	322.1486 / 41.9989	276.0801 / 90.0611	166.128	1219802-74-0
Chloramphenicol-D5 (ISTD)	[M-H] <sup>-</sup>	C <sub>11</sub> H <sub>7</sub> D <sub>5</sub> Cl <sub>2</sub> N <sub>2</sub> O <sub>5</sub>	327.044	326.036	326.0362	-0.195	-0.6	7.93	-	157.066	34.9695	45.9936	261.846	326.034	202480-68-0
Clofianidin-D3 (ISTD)	[M+H] <sup>+</sup>	C <sub>6</sub> H <sub>5</sub> D <sub>3</sub> ClN <sub>5</sub> O <sub>2</sub> S	252.028	253.035	253.0349	0.09	0.4	10.28	-	126.011	90.034	253.0316	-	-	1262776-24-8
m-Toluic acid diethylamide-D7 / DEET-D7 (ISTD)	[M+H] <sup>+</sup>	C <sub>12</sub> H <sub>10</sub> D <sub>7</sub> NO	198.175	199.182	199.1822	-0.036	-0.2	13.3	-	126.094	98.0985	58.9428	199.183	44.0122	1219799-37-7
Diclofenac-D4 (ISTD)	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	C <sub>14</sub> H <sub>7</sub> D <sub>4</sub> NO <sub>2</sub> Cl <sub>2</sub>	299.042	300.049 / 298.035	300.0489 / 298.0343	0.538 / 0.793	0.4 / 0.8	15.7	-	218.0674 / 34.9691	254.0444 / 254.0446	-	-	-	153466-65-0
Diuron-D6 (ISTD)	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	C <sub>8</sub> H <sub>4</sub> D <sub>6</sub> Cl <sub>2</sub> N <sub>2</sub> O	238.054	239.062 / 237.047	239.0619 / 237.0473	1.140351 / -0.2	0.03 / -0.2	13.73	-	78.0823 / 185.9522	52.1028 / 34.9693	239.0608 / 237.0473	- / 41.9985	- / 149.9749	1007536-67-5
Mecoprop-D3 (ISTD)	[M-H] <sup>-</sup>	C <sub>10</sub> D <sub>3</sub> H <sub>6</sub> ClO <sub>3</sub>	217.059	216.051	216.0512	-0.058	-0.3	14.73	-	216.033	34.9694	197.0487	-	-	352431-15-3
Mesotrione-D4 (ISTD)	[M+H] <sup>+</sup> / [M-H] <sup>-</sup>	C <sub>14</sub> H <sub>9</sub> D <sub>4</sub> NO <sub>7</sub> S	343.067	344.074 / 342.059	344.0737 / 342.0594	0.028 / 0.323	0.1 / 0.9	10.55	-	-	-	-	-	-	-

Table-A 6-1 continued.

Metsulfuron-methyl-D3 (ISTD)	[M+H] <sup>+</sup> / [M+H] <sup>-</sup>	C <sub>14</sub> H <sub>12</sub> D <sub>3</sub> N <sub>5</sub> O <sub>6</sub> S	384.093	385.1	385.1003 / 383.0859	-	2.279412	-2.5	11.73	-	170.0759 / 142.0813	135.044	199.0081	209.983	-	-
Salicylic Acid-D4 (ISTD)	[M+H] <sup>-</sup>	C <sub>7</sub> H <sub>2</sub> D <sub>4</sub> O <sub>3</sub>	142.057	141.05	141.0494	-0.08	0.581	-0.6	8.33	-	97.0595	69.0648	44.9981	-	-	78646-17-0
Sulcotriione-D4 (ISTD)	[M+H] <sup>+</sup> / [M+H] <sup>-</sup>	C <sub>14</sub> H <sub>6</sub> D <sub>4</sub> ClO <sub>5</sub> S	332.042	333.050 / 331.035	333.0493 / 331.0348	1.01 / 0.581	0.581	1.0	11.82	-	-	-	-	-	-	-
Tebuconazol-D6 (ISTD)	[M+H] <sup>+</sup>	C <sub>16</sub> H <sub>16</sub> D <sub>6</sub> ClN <sub>3</sub> O	314.19	314.19	314.1903	0.212	0.212	0.7	16.43	-	72.0537	279.232	95.0853	314.182	-	-
Azoxystrobin (QC)	[M+H] <sup>+</sup>	C <sub>22</sub> H <sub>17</sub> N <sub>3</sub> O <sub>5</sub>	403.117	404.124	404.1234	-0.731	-0.731	-1.8	14.5	98.12	372.097	344.103	329.0792	345.106	330.082	131860-33-8
ABC 700* (QC)	[M+H] <sup>+</sup>	C <sub>6</sub> H <sub>6</sub> F <sub>2</sub> N <sub>2</sub> O <sub>2</sub>	176.04	177.047	177.1234	-0.027	-0.027	-0.2	4.98	101.6	137.035	42.0344	109.0399	177.049	-	-
Ciprofloxacin (QC)	[M+H] <sup>+</sup>	C <sub>17</sub> H <sub>18</sub> FN <sub>3</sub> O <sub>3</sub>	331.133	332.14	332.14	-0.54	-0.54	-1.6	7.05	109	314.13	231.057	332.1387	288.15	98.9847	85721-33-1
Clothianidin (QC)	[M+H] <sup>+</sup>	C <sub>8</sub> H <sub>8</sub> ClN <sub>5</sub> O <sub>2</sub> S	249.009	250.016	250.0161	0.111	0.111	0.4	7.79	96.37	131.967	113.017	169.0547	110.072	70.9959	210880-92-5
Fluopyram (QC)	[M+H] <sup>+</sup>	C <sub>16</sub> H <sub>11</sub> ClF <sub>6</sub> N <sub>2</sub> O	396.046	397.054	397.0533	-0.35	-0.35	-0.9	15.31	94.46	173.022	208.014	397.0544	1445.03	130.03	658066-35-4
Imidacloprid (QC)	[M+H] <sup>+</sup>	C <sub>9</sub> H <sub>10</sub> ClN <sub>5</sub> O <sub>2</sub>	255.052	256.06	256.0593	-0.234	-0.234	-0.9	8.32	98.01	175.098	209.059	84.0541	128.027	-	138261-41-3
Lactofen (QC)	[M+NH <sub>4</sub> ] <sup>+</sup>	C <sub>19</sub> H <sub>15</sub> ClF <sub>3</sub> NO <sub>7</sub>	461.049	479.083	479.0823	-0.393	-0.393	-0.8	17.86	103.3	344.998	222.978	-	-	-	77501-63-4
Moxifloxacin (QC)	[M+H] <sup>+</sup>	C <sub>11</sub> H <sub>17</sub> NO <sub>2</sub>	195.126	402.182	402.182	-0.365	-0.365	-0.9	9.15	93.35	402.182	384.166	403.1892	364.162	261.106	151096-09-2
Thiacloprid (QC)	[M+H] <sup>+</sup>	C <sub>10</sub> H <sub>9</sub> ClN <sub>3</sub> S	252.027	253.031	253.0305	-0.454	-0.454	-1.8	10.28	101.6	126.011	90.034	-	-	-	111988-49-9
Acifluorfen (QC)	[M+H] <sup>-</sup>	C <sub>14</sub> H <sub>7</sub> ClF <sub>3</sub> NO <sub>5</sub>	360.996	259.989	359.989	-0.248	-0.248	-0.5	15.36	94.57	194.984	222.03	286.0016	315.999	45.9938	50594-66-6
Benzoic Acid (QC)	[M+H] <sup>-</sup>	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	122.037	121.03	121.0296	0.117	0.117	1.0	5.14	89.34	77.0408	121.027	92.0274	-	-	65-85-0
Benzenesulfonic Acid (QC)	[M+H] <sup>-</sup>	C <sub>6</sub> H <sub>6</sub> O <sub>3</sub> S	158.004	156.996	156.9963	-0.152	-0.152	-1.0	5.14	110	79.9576	93.0347	156.9967	-	-	98-11-3
Salicylic Acid (QC)	[M+H] <sup>-</sup>	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	138.032	137.024	137.0243	-0.137	-0.137	-1.0	8.35	105.2	93.0348	65.0399	-	-	-	69-72-7

### 6.6.1.3 *Different Gradients for Toxicology Determination with HPTLC*

The industrial wastewater sample comprised unknown of interest and the reference material (2-chloro-4,6-dinitrophenol, 3,4-dichloro-2,6-dinitrophenol and 4-chloro-2,6-dinitrophenol) were spotted directly in lines onto high-performance thin-layer chromatography (HPTLC) plates LiChrosphere Silica 60 F254 (Merck KGaA, Darmstadt, Germany). The plates were developed via different gradients (an acid-based gradient, an alkaline-based gradient and an ether-based gradient) using multiple developments on each plate with various solvents. For the acid-based gradient (see Figure-A 6-2) and ether-based gradient (see Figure-A 6-4) the TLC plates were developed in a chamber with formic acid (7 M) saturated ambient air for protonation of the analytes. The mobile phase for acid-based gradient consisted of methanol, acetonitrile, dichloromethane and carbon disulphide and for ether-based gradient tert-butyl methyl ether with acetonitrile (10%) and n-pentane. For the alkaline-based gradient (see Figure-A 6-3) the TLC plates were developed in a chamber with ammonia (3 percent solution) saturated ambient air for deprotonation of the analytes. For mobile phase components methanol, dichloromethane and n-hexane were used. The development with acid-based gradient was done in 25 steps. The order of chromatography is then from polar solvent to nonpolar for which the sample peaks are eluting in the same manner (nonpolar run further, polar components stay at the bottom of plates). UV spectra were measured with the densitometric method (at  $\lambda = 190$  nm,  $\lambda = 200$  nm,  $\lambda = 220$  nm,  $\lambda = 240$  nm,  $\lambda = 260$  nm,  $\lambda = 280$  nm,  $\lambda = 300$  nm). The plates' surfaces were afterwards coated with freshly prepared bacteria solution and tested for active bioluminescence. Bacteria solution of *Aliivibrio fischeri* LCK 484 (Dr. Bruno Lange GmbH & Co KG, Duesseldorf, Germany) was prepared using the LCK 484 Test-Kit to grow in standard LB-Medium. The coated plates (see App. 1d, Table-A 6-2) were measured instantly by long exposure (approx. 3 min) under the camera MicroChemi 4.2 (dnr Bio-imaging Systems Ltd., Neve Yamin, Israel) and determined using 'GelCapture' software (dnr Bio-imaging Systems Ltd., Neve Yamin, Israel).

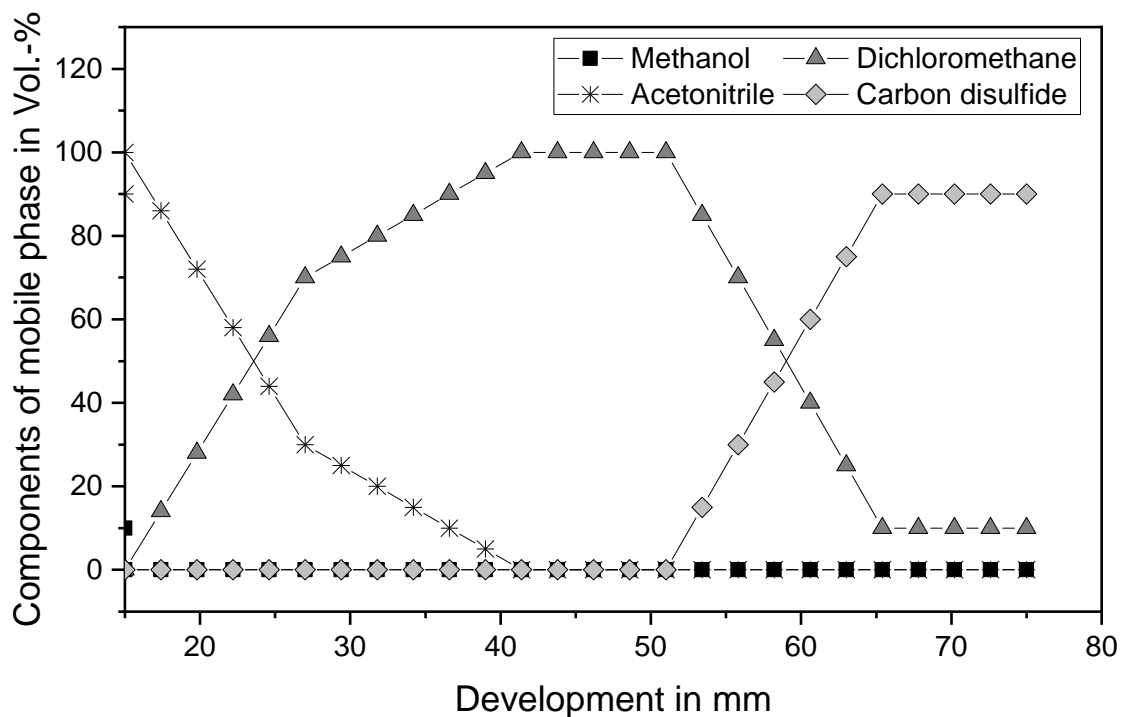


Figure-A 6-2 – Acid-based gradient for development of HPTLC plates.

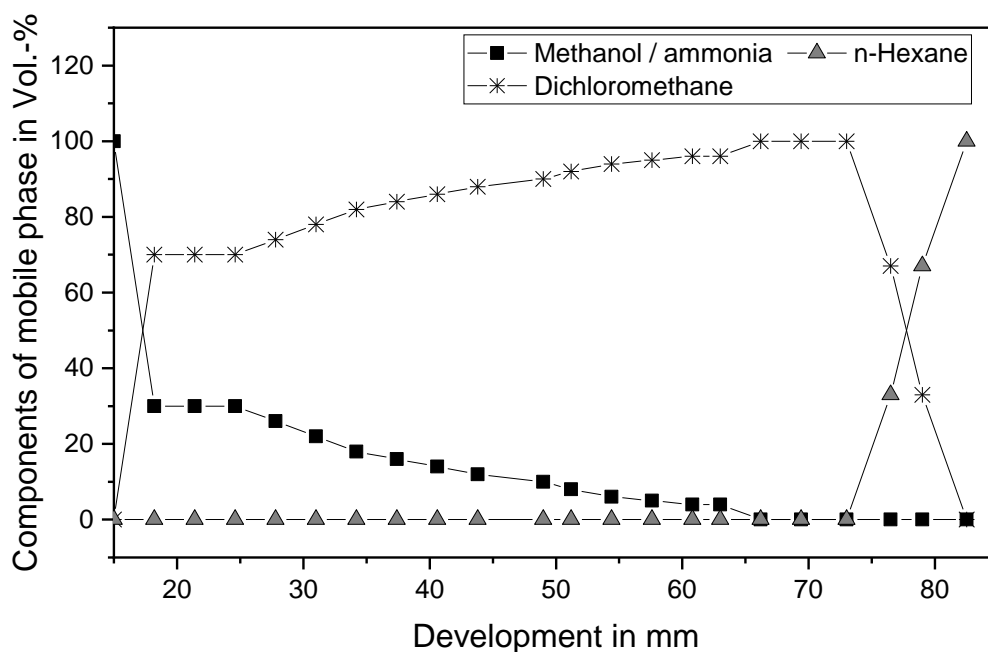


Figure-A 6-3 – Alkaline-based gradient for development of HPTLC plates.

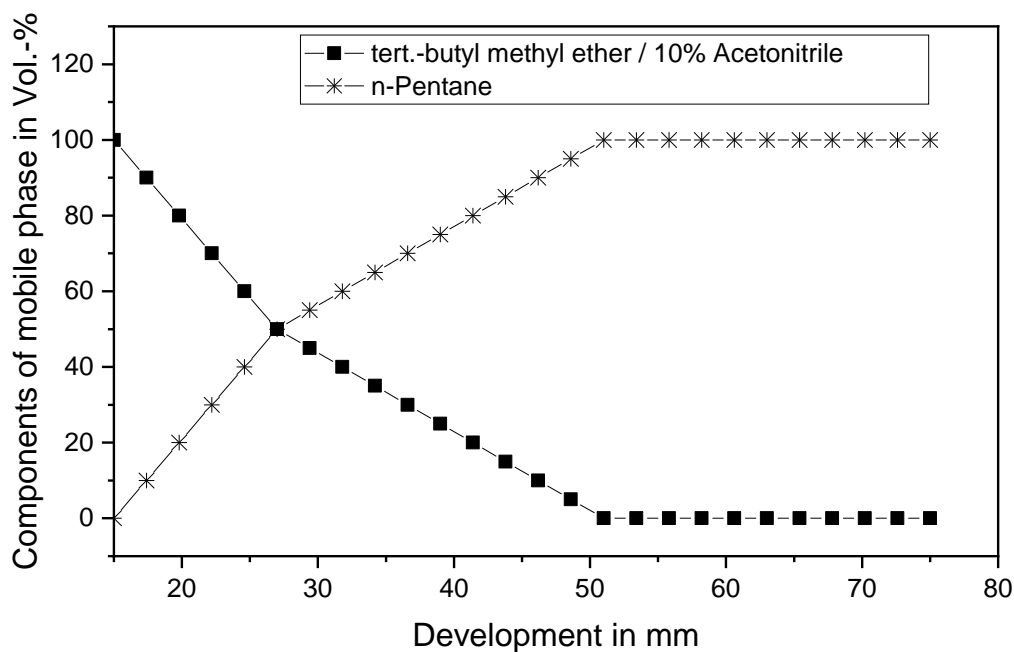


Figure-A 6-4 - Ether-based gradient for development of HPTLC plates.

#### 6.6.1.4 Toxicity test with HPTLC

Table-A 6-2 – Description of schematic order for toxicity testing (Used bacteria: *Aliivibrio fischeri*).

Lane	Applied volume	Applied concentration	Sample
2	3.00 µL	1.00 mL / 10.0 mL*	Wastewater sample I
3	2.00 µL	original	Wastewater sample I
4	5.00 µL	original	Wastewater sample I
5	10.0 µL	10 mg/L	Home-made pos control
6	5.00 µL	original	Wastewater sample II
7	2.00 µL	original	Wastewater sample II
8	3.00 µL	1.00 mL / 10.0 mL*	Wastewater sample II
9	10.0 µL	10 mg/L	3,4-dichloro-2,6-dinitrophenol
10	5.00 µL	10 mg/L	3,4-dichloro-2,6-dinitrophenol
11	2.00 µL	10 mg/L	3,4-dichloro-2,6-dinitrophenol
12	1.00 µL	100 mg/L	2-chloro-4,6-dinitrophenol
13	1.00 µL	100 mg/L	4-chloro-2,6-dinitrophenol
14	1.00 µL	100 mg/L	3,5-dichlorophenol

\*dilution factor of 10

### 6.6.1.5 Signal Cutting for Manual Fractionation

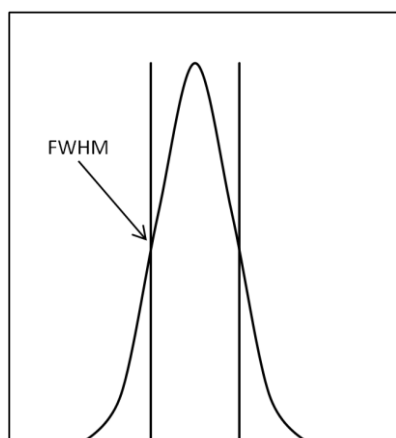


Figure-A 6-5– Representation of ideal signal cutting (dashed lines) at full width at half maximum (FWHM) of normal distributed signal peak.

## 6.6.2 Supplement for Chapter Results and Discussion

### 6.6.2.1 Determination of double bond equivalents

To verify the structure hypothesis, the double bond equivalents (DBE) from the proposed sum formula (with C = carbon atoms, H = hydrogen atoms, N = nitrogen atoms and X = halogens) was calculated, which provided a DBE of six.

$$\begin{aligned} \text{DBE} &= (2 * \text{C} - \text{H} + \text{N} - \text{X} + 2)/2 \\ &= (2 * 6 - 2 + 2 - 2 + 2)/2 \\ &= 6 \end{aligned}$$

A DBE of six means there are six degrees of unsaturation (double bond or a ring system) present in the organic molecule. Thus, five double bonds and one ring, which matches the structure of the suspected dichlorodinitrophenol.

### 6.6.2.2 Constitutional Isomers of Dichlorodinitrophenol

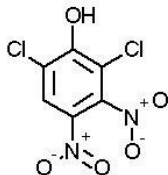
2,3-dichloro-5,6-dinitrophenol



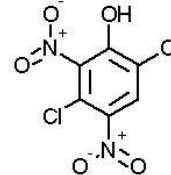
2,3-dichloro-4,5-dinitrophenol



2,6-dichloro-3,4-dinitrophenol



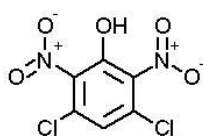
3,6-dichloro-2,4-dinitrophenol



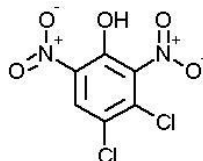
2,6-dichloro-3,5-dinitrophenol



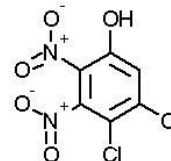
3,5-dichloro-2,6-dinitrophenol



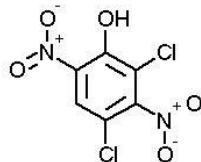
3,4-dichloro-2,6-dinitrophenol



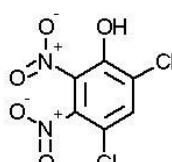
4,5-dichloro-2,3-dinitrophenol



2,4-dichloro-3,6-dinitrophenol



4,6-dichloro-2,3-dinitrophenol



2,3-dichloro-4,6-dinitrophenol

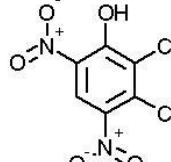


Figure-A 6-6 – Constitutional isomers of structural proposal dichlorodinitrophenol for unknown.

### 6.6.2.3 Toxicity Test with HPTLC

#### Pre-tests:

For assessment of substance toxicity, the following pre-tests were performed. The sample including the proposed compound and the reference material (3,4-dichloro-2,6-dinitrophenol) were analysed by high-performance thin layer chromatography (HPTLC). The silica gel-coated thin layer chromatography (TLC) plate was performed using three different gradients for validation. At the start, at 10 mm there were strongly polar substances that were toxic. In all experiments there was a signal at the same separation front (RF: ~ 70 mm for acid-based gradient and RF: 38 mm for alkaline-based gradient) but showed different UV spectra (Figure-A 6-5 and Figure-A 6-6), with different maxima at different wavelengths. Furthermore, in all experiments there was a front peak at each gradient which was defined the total separation front. The substances could be identified via UV spectrum and relative migration distance. The reference substance in Figure-A 6-5 appeared in the same chromatographic window. The results indicated that the unknown compound had to be a di- or mono-chlorinated dinitrophenol



(amount of chlorine has no influence on the separation in HPTLC) which confirmed the assumption. However, the unknown compound did not appear to be the reference material but might be a superposition of several components of the same chemical structure. It is likely that in the wastewater sample various chlorine-nitro-phenolic compounds were present. Furthermore, the chromatograms showed a high salt and nitrate concentration eluting at the beginning.

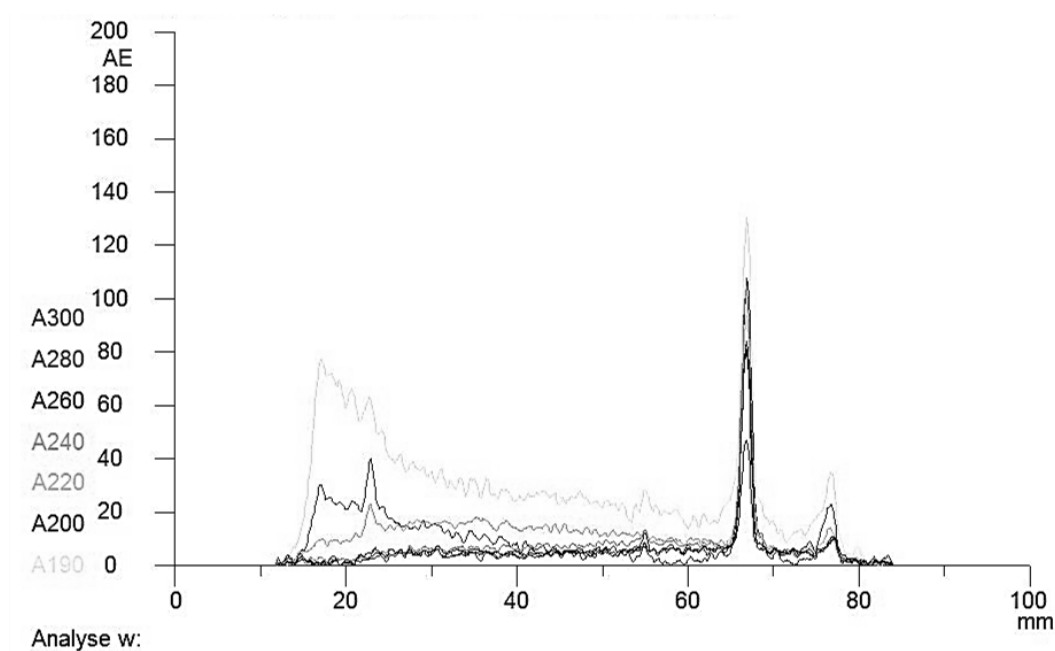


Figure-A 6-7 – Densitogram of reference material 3,4-dichloro-2,6-dinitrophenol (50 ng) measured using acid-based gradient under UV. The colours show the respective wavelengths at which measurements were taken in parallel.

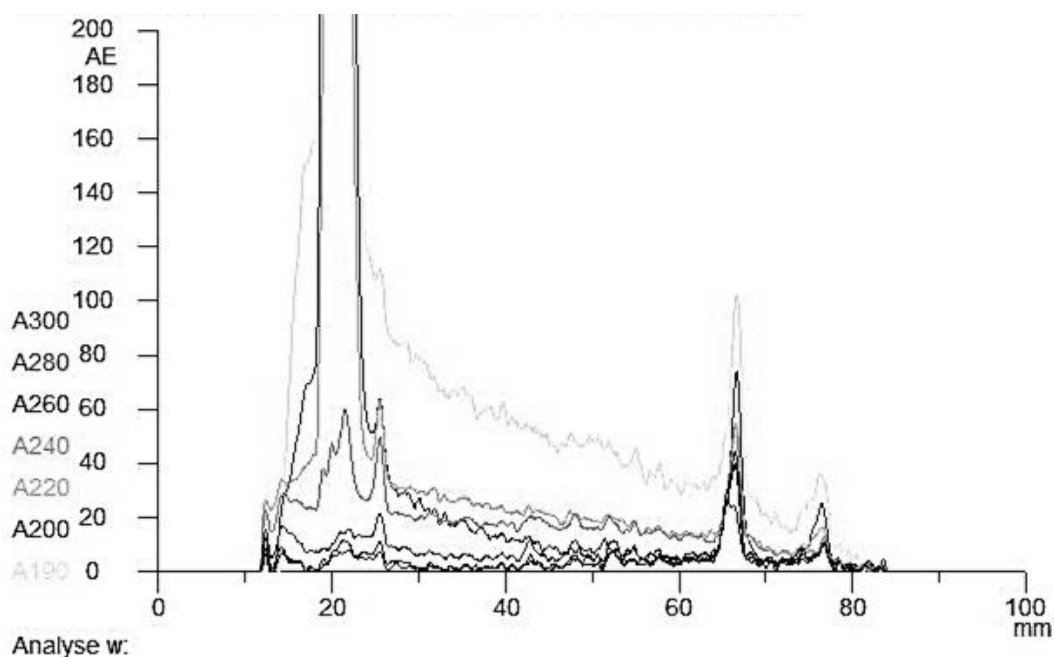


Figure-A 6-8 – Densitogram of unknown substance measured using acid-based gradient under UV. The colours show the respective wavelengths at which measurements were taken in parallel.

### Results:

The substance toxicity was assessed by HPTLC coupled to *Aliivibrio fischeri* bioassay. Pre-tests for HPTLC are given in supporting information (see App. 2c). Black spots indicate bioluminescence quenching and consequently toxicity. 10  $\mu\text{L}$ , 5.0  $\mu\text{L}$  and 2.0  $\mu\text{L}$  of the original sample (prioritisation approach of offline 2D LC is not necessary anymore) was applied on the TLC plate. Even for the 2.0  $\mu\text{L}$  application there was an extinction of luminescence for the component (highest peak with main absorption at 240 nm) at a travel distance of 36 mm using 'alkaline' gradient observed. Based on the safety data sheet of authentic standard, the compound does not need to be labelled in accordance with EC directives or respective national laws (Generic MSDS (EU), Version 3.8, 2015-02-06, ENAMINE Ltd. Kyiv, Ukraine). In the European Chemical Agency (ECHA) database no results were found for dichlorodinitrophenols. However, some 1960's patents say that some chloronitrophenol pesticides may be toxic (Pyne, 1963, 1962).

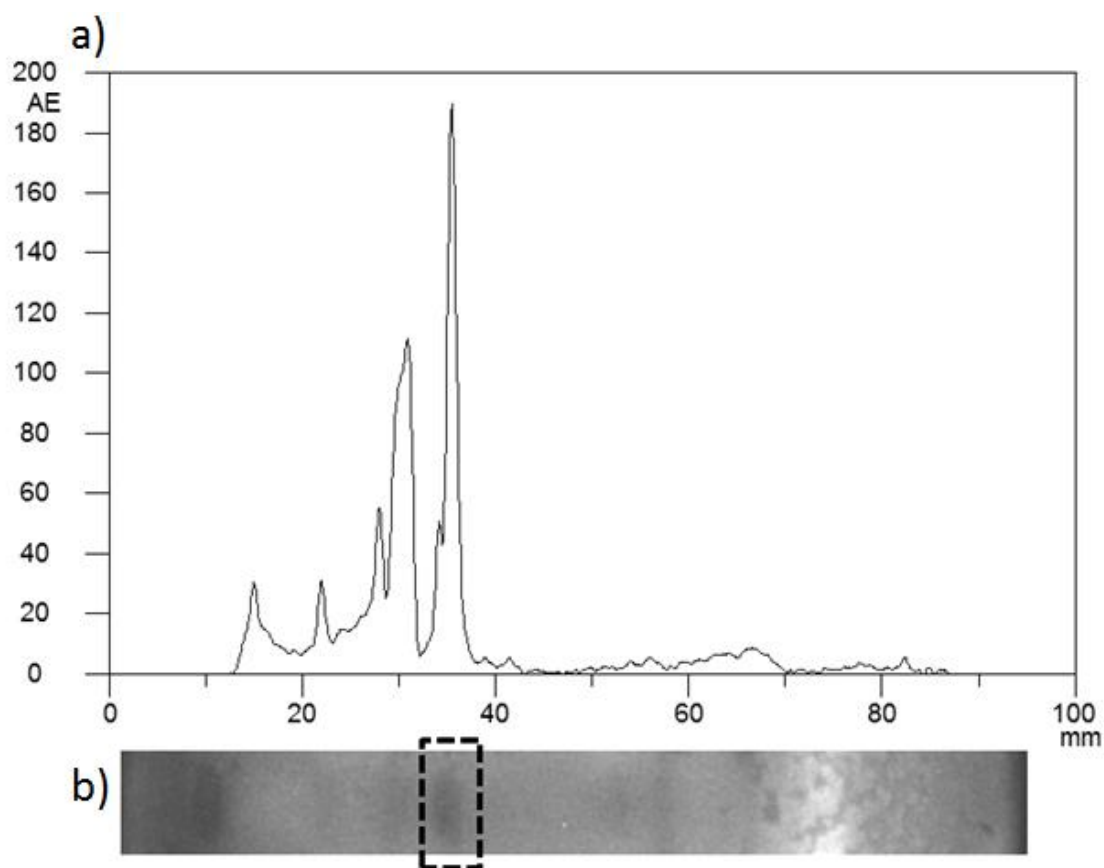


Figure-A 6-9 – Assessment of substance toxicity by HPTLC. In a) the densitogram (recorded at 240 nm) and in b) the toxicity result of eluent fraction of wastewater sample are recorded using 'alkaline' gradient (see App. 1c, Figure-A 6-2 to Figure-A 6-4). The black box highlights the black spots indicating bioluminescence quenching and consequently toxicity of peak of interest. The placement of TLC plate is listed in App. 1d, Table-A 6-2.

### 6.6.2.4 Results of Quantification Analysis

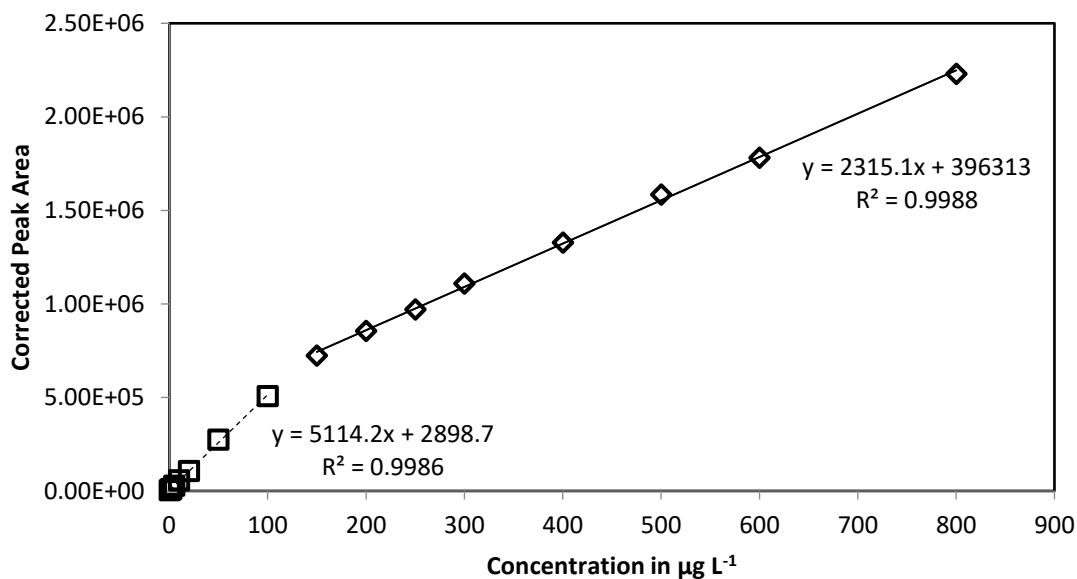


Figure-A 6-10 – Calibration curve of lower concentration range and higher concentration range in mobile phase A of LC-HRMS (Milli-Q® with 0.1% formic acid) of reference material 3,4-dichloro-2,6-dinitrophenol for quantification of isolated material using LC-UV.

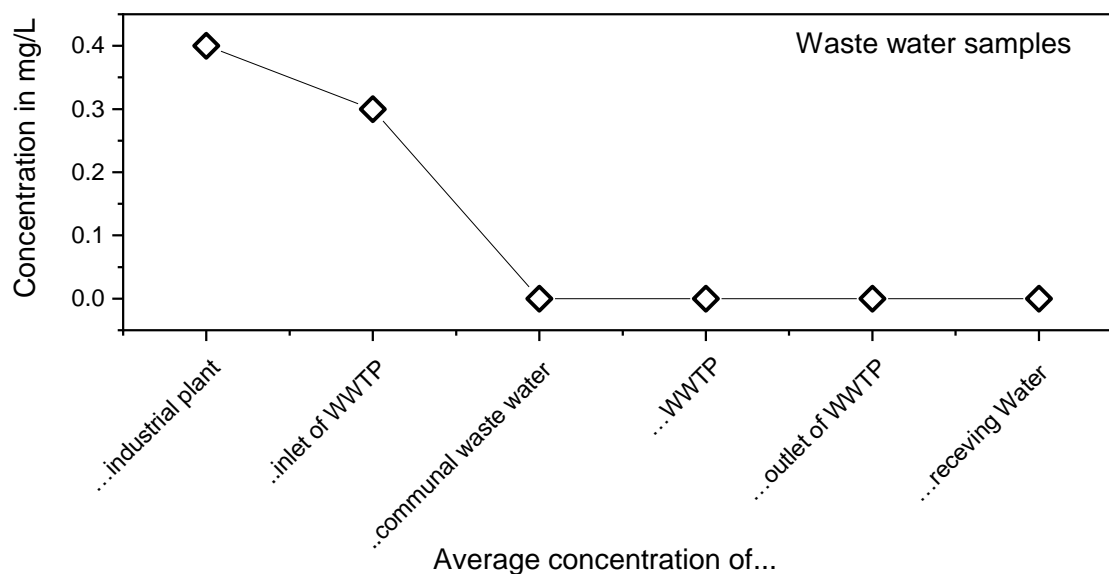


Figure-A 6-11 – Determination of average quantified (by authentic standard 3,4-dichloro-2,6-dinitrophenol in Milli-Q® water) concentration of identified 'known unknown' in wastewater samples of industrial plant and the WWTP in mg L<sup>-1</sup>.

### 6.6.3 References

Loos, M., 2018. enviMass. Zenodo.

## Chapter 7 General Conclusions and Outlook

In this thesis, a novel non-target screening (NTS) approach by high-resolution mass spectrometry coupled to liquid chromatography (LC-HRMS) for the expanded monitoring of industrial wastewater was developed. The results of this thesis demonstrate that the prioritisation is a necessary step for the identification of unknown features that may be trace organic compounds (TrOCs) in industrial wastewater. In detail, several technical (chapter 6) and statistical (chapter 4, chapter 5) prioritisation workflows are presented. The prioritisation providing candidates for structure elucidation was a key point. On this basis, the assessment of an industrial wastewater treatment plant (WWTP) got possible.

In this thesis, NTS was based on reversed-phase liquid chromatography with time-of-flight mass spectrometry and electrospray ionisation (LC-ESI-qTOF-MS), which describes the standard procedure in the detection of thermally labile or low-volatility, (slightly) polar TrOCs in water (Bader et al., 2016b; Blum et al., 2017; Krauss et al., 2010; Schymanski et al., 2014b). In fact, the polarity range was expanded for monitoring TrOCs during this thesis, but the analysis does still cover only a part of TrOCs in wastewater. To expand the analytical window, regarding more polar or less polar compounds, other LC phases would allow separation and detection of these compounds not captured by reversed-phase LC. For example, hydrophilic interaction liquid chromatography (HILIC) (Zahn et al., 2016) and ion chromatography (IC) (Agostini, 2018; Gallidabino et al., 2018) are designed for more polar compounds like transformation products, metabolites or degradation products. Gas chromatography (GC) can be used for less polar compounds like polychlorinated biphenyls (PCBs), brominated diphenyl ethers (BDEs) and polycyclic aromatic hydrocarbons (PAHs), which are potential components of industrial wastewater (Domínguez et al., 2018, 2017). By combining the LC-HRMS method with GC and HILIC/IC, the analysis of industrial wastewater could be further expanded and ensure more holistic monitoring.

Even when the technical requirements regarding the analysis systems become feasible, the need for optimising the data evaluation workflows remains before NTS can be implemented in routine. For example, the needed evaluation time for 24-h composite flow wastewater samples of one week ( $N = 7$ ) is excessively long. Therefore, optimisation needs to be performed before a routine application. The data analysis of these samples could take up to weeks for evaluation. For NTS, many adjustments could be assigned for optimising the evaluation effort, e.g. the reduction of false positives and the development of individual spectral libraries of production plants in the industrial area. Furthermore, NTS should be based on standardised workflows analysing wastewater samples and above all, proper documentation should be required.

As data processing is the most crucial step in the NTS workflow, detailed documentation on the workflow including implementation of filtering steps, all parameter settings and criteria should be mandatory in all future studies (Hohrenk et al., 2020). Furthermore, the NTS workflow and time window of measurement after sampling should be adapted for every sample to reduce time in method development and reduce systematic errors. This will lead to a further positive effect by allowing for comparable results during routine monitoring. Comparability is particularly crucial for retrospective analysis, which was already performed in other studies (Alygizakis et al., 2018; Creusot et al., 2020; Polgár et al., 2012). For quality assurance, operating figures of the used HRMS, e.g. detector voltages, as well as troubleshooting and cleaning, should be documented. These data serve as additional information and might be necessary for interpretation of data regarding, e.g., intensity differences. Expanding data libraries of sampling sites of production plants will ensure fingerprints. This was already shown as an initial study in chapter 5. Libraries containing fingerprints of every production cycle of corresponding production plants could accelerate data analysis by being aware of normal wastewater composition. The fingerprints could be designed after repeated measurements, storing each recurring high-intensive feature, with its spectrum as well as the sampling site's specific background signals in the production-specific database. On this basis, contaminations causing incidents in a WWTP, e.g. foam formation, will almost certainly be identified by analysing differences in the production-specific feature fingerprint. Furthermore, an application of NTS in online measurement becomes realisable by monitoring, for example, the rarity score described in chapter 5. At least, reduction of false positives will accelerate data processing which is still a problem (Bader et al., 2016a; Nürenberg et al., 2015). Bader et al. as well as Nürenberg et al. already assigned the different number of false positives by applying replication measurements. By measuring, e.g., triplicates, and considering only features, which occur in all triplicates, false positives were reduced. For this thesis, the application of measuring triplicates would be useful, e.g., in chapter 6 in the second chromatographic separation dimension prioritising the relevant feature. In other approaches, e.g., in the trend analysis prioritisation of long-term time series (chapter 4 and chapter 5), the measurement of replicates is unnecessary.

Applying NTS in routine monitoring, proper data storage is a further necessity. On the one hand, data storage enables retrospective analysis and on the other hand, the stored data can serve as retaining samples for quality assurance. Regarding data archiving, access to data is the limiting factor because of data processing and data conformity problems. Consequently, a filing space is required that enables high storage capacities and allows later reanalysis over several years (depending on quality assurance).

For the identification of unknowns in several studies, comparisons with online databases led to successful identification for prioritised features (Gago-Ferrero et al., 2015; Schymanski et al., 2014b). The molecular formulae proposed can be searched for in commercial databases, trying to assign a structure. In chapter 4, N-methylpyrrolidone was elucidated for one prioritised feature by *ChemSpider* (Royal Society of Chemistry, 2017) and subsequently confirmed by reference material. As NTS studies have become more widespread in environmental, pharmaceutical and food analysis, as well as numerous –omics (e.g. metabolomics) approaches, forensic investigations and so on, the quality assurance of spectral libraries has become more important (Milman and Zhurkovich, 2016). However, common mass spectral libraries do not comprise a wide range of industrial wastewater contaminants. Furthermore, a challenge of the MS/MS spectral database has been the variability in fragmentation reactions caused by limited standardisation and harmonisation of experimental conditions (Oberacher et al., 2020). A further complication is that databases were established on different instruments (i.e., qTOF and various Orbitrap hybrid instruments). Therefore, there is a high demand for expanding and improving open source databases, especially for industrial chemicals. Besides, information on instrument parameters, e.g. collision energies, should be implemented to ensure comparability. To compensate the lack of TrOCs in open source databases, chemical databases were built in the past that could be used with identification tools such as *in silico* fragments, to expand the chemical space which is available for investigation (Ruttkies et al., 2016). In chapter 6, a dichlorodinitrophenol was elucidated using, e.g. *MetFrag*. Nevertheless, entirely unknown chemicals, which were by now not reported, cannot be found in such databases. In these cases, further preliminary information utilising complementary techniques to HRMS, e.g. infrared/Raman spectroscopy or nuclear magnetic resonance spectroscopy (NMR), as  $^1\text{H-NMR}$  or  $^{13}\text{C-NMR}$ , need to be performed. In chapter 6 and other studies of the literature, NMR spectroscopy proved to be feasible (Armbruster et al., 2015; De Vijlder et al., 2018; Purschke et al., 2020; Wick et al., 2011). However, because of partially low concentrations of TrOCs in wastewater samples, NMR analysis becomes insufficient and enrichment procedures are needed, which are also time consuming. Therefore, NMR will hopefully become more sensitive in the future, allowing its more widespread use as a complementary technique for the structure elucidation of compounds at environmental concentrations.

To overcome the mentioned problems of missing standardisation in NTS workflow and fragmentation procedures, recently some optional harmonisation workflows were introduced. The 'Non-Target Screening' expert committee of the German Water Chemistry Society published a guideline on the use of NTS using LC-ESI-HRMS in water analysis in 2019 ("Non-Target Screening" expert committee of the German Water Chemistry Society, 2019). The guideline introduces fundamental aspects in the use of LC-HRMS devices, information on

potential contamination in sampling and measurements, but also aspects of data evaluation and quality assurance measures. Additionally, to guarantee comparability of different laboratories, a uniform categorisation was introduced for the identification of unknown substances by NTS (Schymanski et al., 2014a). Until now, there are no regulations regarding the implementation of NTS in routine water quality assessment. Due to the introduction of the NTS-guideline (2019), the application of NTS in water analysis will grow, which will sooner or later enforce harmonisations and regulations (“Non-Target Screening” expert committee of the German Water Chemistry Society, 2019).

Furthermore, for application in routine analysis, the introduced time trend prioritisation method of chapter 4, could be expanded regarding real-time monitoring of WWTPs. Machine learning concepts could realise on-line time-related analysis of changes in the WWTP. Therefore, for example, features causing adverse effects on nitrification in the WWTP could be monitored, resulting in an early-warning system to protect the proper functioning of the WWTP. Online LC-HRMS has already been applied for industrial wastewater (Wortberg and Kurz, 2019). However, at first, the workflow needs to be standardised and some further validation experiments going beyond the proof-of-concept should be performed.

Finally, additional prioritisation methods of NTS data focusing on other properties for selecting features should be implemented. For example, prioritisation methods focusing on selecting unknowns based on their potential toxic effect have been developed. Applying these methods on wastewater unknowns, the identification effort can focus only on those compounds that may pose a risk to the environment (unknowns found in WWTP effluents) or the biological treatment step of the WWTP (unknowns found in the WWTP influent). Therefore, the biological purification in the WWTP will be optimised by determining substances with toxic effects in the wastewater, for which a complete identification is not even necessary. These prioritisation strategies are based on effect-direct analysis (EDA) (Muschket et al., 2018; Stütz et al., 2019; Tousova et al., 2018). Samples are first tested with *in vitro* and/or *in vivo* bioassays to determine if there is a potential toxic effect. In case of a positive result, a fractionation is applied to isolate these fractions of the sample with a toxic effect. The results of chapter 6 demonstrated initial tests of the coupling of biological effects, data analysis and structure elucidation. In future studies, this point will be of great interest because of the continuing growth in the protection of the environment and human health (Altenburger et al., 2019).

In conclusion, this thesis has shown that NTS applying different data analysis strategies is a meaningful way to characterise and prioritise potential TrOCs occurring in industrial wastewaters. The application in routine analysis has high potential expanding the monitoring of wastewater and sewage system. Especially when data processing will be accelerated, the



data storage be ensured and the awareness of the limits of the detectable analytes by only using LC-HRMS be raised, NTS will improve water analysis. Additionally, the application of the presented concepts is not limited to industrial wastewaters and may be transferred to other research fields where monitoring over space and time is conducted, prioritisation of relevant features for identification is in demand and a concept for process assessment is required. For example, the prioritisation methods could be useful in biomonitoring studies for future biomarker searches or in forensics, tracking synthetic drug laboratories by sewage monitoring.

### 7.1 References

- Agostini, V., 2018. Evolution of Laboratory Procedures for Water Quality Analysis, in: Gilardoni, A. (Ed.), *The Italian Water Industry: Cases of Excellence*. Springer International Publishing, Cham, pp. 119–128. [https://doi.org/10.1007/978-3-319-71336-6\\_8](https://doi.org/10.1007/978-3-319-71336-6_8)
- Altenburger, R., Brack, W., Burgess, R.M., Busch, W., Escher, B.I., Focks, A., Mark Hewitt, L., Jacobsen, B.N., de Alda, M.L., Ait-Aissa, S., Backhaus, T., Ginebreda, A., Hilscherová, K., Hollender, J., Hollert, H., Neale, P.A., Schulze, T., Schymanski, E.L., Teodorovic, I., Tindall, A.J., de Aragão Umbuzeiro, G., Vrana, B., Zonja, B., Krauss, M., 2019. Future water quality monitoring: improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. *Environ. Sci. Eur.* 31, 1–17. <https://doi.org/10.1186/s12302-019-0193-1>
- Alygizakis, N.A., Samanipour, S., Hollender, J., Ibáñez, M., Kaserzon, S., Kokkali, V., Van Leerdam, J.A., Mueller, J.F., Pijnappels, M., Reid, M.J., Schymanski, E.L., Slobodnik, J., Thomaidis, N.S., Thomas, K. V., 2018. Exploring the Potential of a Global Emerging Contaminant Early Warning Network through the Use of Retrospective Suspect Screening with High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 52, 5135–5144. <https://doi.org/10.1021/acs.est.8b00365>
- Armbruster, D., Happel, O., Scheurer, M., Harms, K., Schmidt, T.C., Brauch, H.-J., 2015. Emerging nitrogenous disinfection byproducts: Transformation of the antidiabetic drug metformin during chlorine disinfection of water. *Water Res.* 79, 104–118. <https://doi.org/https://doi.org/10.1016/j.watres.2015.04.020>
- Bader, T., Schulz, W., Kümmerer, K., Winzenbacher, R., 2016a. General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Anal. Chim. Acta* 935, 173–186. <https://doi.org/10.1016/j.aca.2016.06.030>
- Bader, T., Schulz, W., Lucke, T., 2016b. Application of non-target analysis with LC-HRMS for the monitoring of raw and potable water: Strategy and results, in: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening – Strategies and Workflows Volume 2*. ACS Symposium Series, Washington, DC., pp. 49–70. <https://doi.org/10.1021/bk-2016-1242.ch003>
- Blum, K.M., Andersson, P.L., Renman, G., Ahrens, L., Gros, M., Wiberg, K., Haglund, P., 2017. Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. *Sci. Total Environ.* 575, 265–275. <https://doi.org/10.1016/j.scitotenv.2016.09.135>
- Creusot, N., Casado-Martinez, C., Chiaia-Hernández, A., Ferrari, B.J.D., Fischer, S., Fu, Q., Munz, N., Singer, H., Stamm, C., Tlili, A., Hollender, J., 2020. Retrospective Screening of High Resolution Mass Spectrometry Archived Digital Samples can

- Improve Environmental Risk Assessment of Emerging Contaminants: A Case Study on Antifungal Azoles (submitted). *Environ. Int. J.* 139, 105708. <https://doi.org/10.1016/j.envint.2020.105708>
- De Vijlder, T., Valkenburg, D., Lemièrre, F., Romijn, E.P., Laukens, K., Cuyckens, F., 2018. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* 37, 607–629. <https://doi.org/10.1002/mas.21551>
- Domínguez, I., Arrebola, F.J., Gavara, R., Martínez Vidal, J.L., Frenich, A.G., 2018. Automated and simultaneous determination of priority substances and polychlorinated biphenyls in wastewater using headspace solid phase microextraction and high resolution mass spectrometry. *Anal. Chim. Acta* 1002, 39–49. <https://doi.org/10.1016/j.aca.2017.11.056>
- Domínguez, I., Arrebola, F.J., Romero-González, R., Nieto-García, A., Martínez Vidal, J.L., Garrido Frenich, A., 2017. Solid phase microextraction and gas chromatography coupled to magnetic sector high resolution mass spectrometry for the ultra-trace determination of contaminants in surface water. *J. Chromatogr. A* 1518, 15–24. <https://doi.org/10.1016/j.chroma.2017.08.061>
- Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* 49, 12333–12341. <https://doi.org/10.1021/acs.est.5b03454>
- Gallidabino, M.D., Hamdan, L., Murphy, B., Barron, L.P., 2018. Suspect screening of halogenated carboxylic acids in drinking water using ion exchange chromatography – high resolution (Orbitrap) mass spectrometry (IC-HRMS). *Talanta* 178, 57–68. <https://doi.org/https://doi.org/10.1016/j.talanta.2017.08.092>
- Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Anal. Chem.* 92, 1898–1907. <https://doi.org/https://doi.org/10.1021/acs.analchem.9b04095>
- Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: From target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–951. <https://doi.org/10.1007/s00216-010-3608-9>
- Milman, B.L., Zhurkovich, I.K., 2016. Mass spectral libraries: A statistical review of the visible use. *TrAC - Trends Anal. Chem.* 80, 636–640. <https://doi.org/10.1016/j.trac.2016.04.024>
- Muschket, M., Di Paolo, C., Tindall, A.J., Touak, G., Phan, A., Krauss, M., Kirchner, K., Seiler, T.-B., Hollert, H., Brack, W., 2018. Identification of Unknown Antiandrogenic Compounds in Surface Waters by Effect-Directed Analysis (EDA) Using a Parallel Fractionation Approach. *Environ. Sci. Technol.* 52, 288–297. <https://doi.org/10.1021/acs.est.7b04994>
- “Non-Target Screening” expert committee of the German Water Chemistry Society, 2019. Use of non-target screening by means of LC-ESI-HRMS in water analysis. URL <https://www.wasserchemische-gesellschaft.de/de/publikationen/seiten/publikationen-der-wasserchemischen-gesellschaft>
- Nürnberg, G., Schulz, M., Kunkel, U., Ternes, T.A., 2015. Development and validation of a generic nontarget method based on LC-HRMS analysis for the evaluation of different wastewater treatment options. *J. Chromatogr. A* 1426, 77–90. <https://doi.org/10.1016/j.chroma.2015.11.014>

- Oberacher, H., Sasse, M., Antignac, J.P., Guitton, Y., Debrauwer, L., Jamin, E.L., Schulze, T., Krauss, M., Covaci, A., Casero, N.C., Rousseau, K., Damont, A., Fenaille, F., Lamoree, M., Schymanski, E.L., 2020. A European proposal for quality control and quality assurance of tandem mass spectral libraries. *Environ. Sci. Eur.* <https://doi.org/10.1186/s12302-020-00314-9>
- Polgár, L., García-Reyes, J.F., Fodor, P., Gyepes, A., Dernovics, M., Abrankó, L., Gilbert-López, B., Molina-Díaz, A., 2012. Retrospective screening of relevant pesticide metabolites in food using liquid chromatography high resolution mass spectrometry and accurate-mass databases of parent molecules and diagnostic fragment ions. *J. Chromatogr. A* 1249, 83–91. <https://doi.org/10.1016/j.chroma.2012.05.097>
- Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Sci. Total Environ.* 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>
- Royal Society of Chemistry, 2017. ChemSpider [WWW Document]. URL <https://www.chemspider.com/> (accessed 8.5.17).
- Rutkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* 8, 1–16. <https://doi.org/10.1186/s13321-016-0115-9>
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014a. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. <https://doi.org/10.1021/es5002105>
- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014b. Strategies to characterize polar organic contamination in wastewater: Exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol.* 48, 1811–1818. <https://doi.org/10.1021/es4044374>
- Stütz, L., Leitner, P., Schulz, W., Winzbacher, R., 2019. Identification of genotoxic transformation products by effect-directed analysis with high-performance thin-layer chromatography and non-target screening. *J. Planar Chromatogr. - Mod. TLC* 32. <https://doi.org/https://doi.org/10.1556/1006.2019.32.3.1>
- Tousova, Z., Froment, J., Oswald, P., Slobodník, J., Hilscherova, K., Thomas, K. V, Tollefsen, K.E., Reid, M., Langford, K., Blaha, L., 2018. Identification of algal growth inhibitors in treated waste water using effect-directed analysis based on non-target screening techniques. *J. Hazard. Mater.* 258, 494–502. <https://doi.org/10.1016/j.jhazmat.2018.05.031>
- Wick, A., Wagner, M., Ternes, T.A., 2011. Elucidation of the Transformation Pathway of the Opium Alkaloid Codeine in Biological Wastewater Treatment. *Environ. Sci. Technol.* 45, 33374–3385. <https://doi.org/https://doi.org/10.1021/es103489x>
- Wortberg, M., Kurz, J., 2019. Analytics 4.0: Online wastewater monitoring by GC and HPLC. *Anal. Bioanal. Chem.* 411, 6783–6790. <https://doi.org/10.1007/s00216-019-02065-w>
- Zahn, D., Frömel, T., Knepper, T.P., 2016. Halogenated methanesulfonic acids: A new class of organic micropollutants in the water cycle. *Water Res.* 101, 292–299. <https://doi.org/https://doi.org/10.1016/j.watres.2016.05.082>

## Chapter 8 Appendix

### 8.1 List of Abbreviations

[M+H] <sup>+</sup>	Protonated molecule
[M-H] <sup>-</sup>	Deprotonated molecule
2D	Two-dimensional
ABC 700	Industrial chemical 1
Abl.	Sampling in the WWTP
AF	Asymmetric factor
AMTT	2-Amino-4-methoxy-6-trifluoromethyl-1,3,5-triazine
APCI	Atmospheric pressure chemical ionisation
API	Atmospheric pressure ionisation
B	Mobile phase B (organic)
BDE	Brominated diphenyl
C18	Octadecylsilane
CAS	Chemical abstract service
CCF	Count conversion factor
CCIM	4-Chlor-2-cyano-5-(4-methylphenyl)imidazole
CE	Collision energy
CES	Collision energy spread
Da	Dalton
DBE	Double bond equivalent

DDA	Data-dependent acquisition analysis
DEET	N,N-Diethyl-3-methylbenzamide
DEF	Industrial chemical 2
DIA	Data-independent acquisition analysis (MS/MS <sup>all</sup> )
Diglyme	Diethyleneglykoldimethylether
DMSA	N,N-Dimethyl-N'-phenylsulfamid
EDA	Effect-directed analysis
EF	Elution fraction
ESI	Electrospray-ionisation
F	Feature
f <sub>c</sub>	Fold change
FID	Flame ionisation detection
FN	False negatives
FP	False positives
FT-ICR	Fourier-transform ion cyclotron resonance mass spectrometry
FWHM	Full width at half maximum
GC	Gas chromatography
GIA	Group identification algorithm
GPC	Group-wise principal component
GPCA	Group-wise principal component analysis
HCA	Hierarchical cluster analysis
HDPE	High-density polyethylene

HILIC	Hydrophilic interaction liquid chromatography
HPLC	High-performance liquid chromatography
HPTLC	High-performance thin layer chromatography
HR	High-resolution
HRMS	High Resolution Mass Spectrometry
IC	Ion chromatography
ICP	Inductively coupled plasma
ID	Identifier
IDA	Information Dependent Acquisition (SCIEX)
Iminopyrazole Acid-3	1-(3'-Sulfophenyl)-3-methyl-5-aminopyrazole
ISTD	Internal standard of isotope-labelled (deuterated) compounds
ITC	Ion transmission control
<i>k</i>	Retention time factor
LC	Liquid Chromatography
LOD	Limit of detection
LOQ	Limit of quantification
LRMS	Low-resolution mass spectrometry
M	Molecular weight
<i>m/z</i>	Mass-to-charge ratio
MA	Outflow of the WWTP
mDa	Milli Dalton
MEBA	Multivariate empirical bayes approach

MEDA	Missing-data for exploratory data analysis
Melamine	2.4.6-Triamino-1.3.5-triazine
MRM	Multi reaction monitoring
MRM <sup>HR</sup>	Multiple Reaction Monitoring with high resolution
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
MV	MarkerView
N.D.	Not detected
NMR	Nuclear magnetic resonance (spectroscopy)
No.	Number
NP	Normal-phase
NTS	Non-target screening
OES	Optical emission spectrometry
PAH	Polycyclic aromatic hydrocarbon
PC	Principal component
PCA	Principal component analysis
PCB	Polychlorinated biphenyl
pH	' <i>potentia Hydrogenii</i> '; negative decimal logarithm of the hydrogen ion activity
PLS-DA	Partial least square linear-discriminant analysis
ppm	Parts per million
Q	Quadrupole
QC	Quality control

QQQ	Triple quadrupole
qTOF	Quadrupole-time-of-flight
R	Resolution
Rec.	Recovery
Ref.	Authentic standard
Rep.	Replicates
RP	Reversed-phase
rpm	Rounds per minute
rpm	Rounds per minutes
RRT	Relative retention time
RSD	Relative standard deviation
RT	Retention time
S/N	Signal-to-noise ratio
SD	Standard deviation
SP	Sampling site
SWATH <sup>®</sup>	Sequential Window Acquisition of all theoretical Ions (SCIEX)
TBSA	2-(Trifluormethyl)benzenesulfonamide
Tetraglyme	Tetraethyleneglycoldimethylether
TIC	Total ion chromatogram
TMS	Trimethylsilyl (functional group)
TN	True negatives
TOC	Total organic carbon



TOF	Time-of-flight
tp	Time point
TP	True positives
Triglyme	Triethyleneglycoldimethylether
TrOC	Trace organic compounds
UPW	Ultra-pure water
UV	Ultra-violet detection
vDIA	Variable data-independent acquisition
WE	Inflow of the WWTP
WWTP	Wastewater treatment plant
XIC	Extracted ion chromatogram
y	Threshold for GIA

## 8.2 List of Figures and Figures-A

- Figure 1-1 – Components of LC-MS. The LC separates organic molecules. In the ionisation source ions are produced, which are separated according to their mass-to-charge ratio ( $m/z$ ) in the mass analyser. The detector collects the separated ions and converts them into electrical signals. Finally, the data system processes the signals resulting in three-dimensional data of  $m/z$ , retention time and intensity (features (Bader et al., 2016))... - 15 -
- Figure 1-2 – Illustration of mass spectrometry terms on the example of the herbicide diuron. Listed  $m/z$  values correspond to the protonated molecule  $[M+H]^+$ . The mass spectrum shows the distinct isotopic pattern of two chlorine atoms of the structure:  $C_9H_{10}^{35}Cl_2N_2O \triangleq$  nominal mass of 233 (100%),  $C_9H_{10}Cl^{35}Cl^{37}N_2O \triangleq$  nominal mass of 235 (64.71%),  $C_9H_{10}Cl^{37}N_2O \triangleq$  nominal mass of 237 (10.70%). The other occurring ions describe the  $^{13}C$ -isotopes (1% of  $^{12}C$  isotopes). ..... - 18 -
- Figure 1-3 – Schematic overview of the x500R qTOF instrument from SCIEX used in this work. © 2017 AB SCIEX..... - 20 -
- Figure 1-4 – Schematic representation of data evaluation strategies used for LC-HRMS measurements including target screening (reference standards are available), suspected-target screening (prior structural information of the suspects available, but no reference standards are available) and NTS (no previous knowledge and no reference standards are available) based on (Krauss et al., 2010) and (Bletsou et al., 2015)..... - 21 -
- Figure 1-5 – Identification confidence levels in HRMS analysis. Redrafted from Schymanski et al. (Schymanski et al., 2014a)..... - 23 -
- Figure 2-1 – Graphical overview of the content of this thesis. .... - 33 -
- Figure 3-1 – Schematic representation of sampling sites (SP) of used samples. .... - 38 -
- Figure 3-2 – Summary of the LC-HRMS method development including data processing and corresponding validation. The samples used for single-steps are mentioned. The performed tests are listed..... - 41 -
- Figure 3-3 – Result of centrifugation experiments of 17 inflow samples of the WWTP (SP 3), 12 pre-treated samples within the WWTP (SP 4) and 16 samples of outflow of the WWTP (SP 5). The detailed results are shown in Appendix Table-A 3-6. The relative standard deviations (RSD) shown, describe the difference between an initial quantification measurement of spiked samples without centrifugation and the result of suspected-target screening analysis of the centrifuged samples. .... - 42 -
- Figure 3-4 – Recovery results of storage experiments. The samples of three exemplary matrices: WWTP influent samples (SP 3), samples within the WWTP (SP 4) and WWTP effluents (SP 5) were stored for one week in the freezer at  $-22\text{ }^\circ\text{C}$  (a) and in the fridge at  $8\text{ }^\circ\text{C}$  (b). The histograms show the number of analytes recovered with 80 to 120%, 60 to 80%, 40 to 50%, 10 to 40% and less than 10%. The label of the bars indicates the number of samples. The results represent means of nine measurements per matrix. Detailed results are shown in 3.6.2.3, Table-A 3-7. .... - 44 -
- Figure 3-5 – Median of retention time deviation in min of triplicate measurement of all spiked suspected-targets and ISTDs over absolute retention time in min. The error bars represent the upper (maximum of the third quartile, 75<sup>th</sup> percentile) and lower (minimum of first quartile, 25<sup>th</sup> percentile) whiskers with respectively  $N= 3$ . Detailed results are listed in 3.6.2.8, Table-A 3-18. .... - 48 -

- Figure 3-6 –The figure illustrates the median of mass error of spiked suspected-targets and ISTDs in ppm of all measured samples for validation in positive and negative ionisation mode over three months. The error bars represent the upper (maximum of the third quartile, 75<sup>th</sup> percentile) and lower (minimum of first quartile, 25<sup>th</sup> percentile) whiskers. Detailed results are listed in 3.6.2.8, Table-A 3-18..... - 49 -
- Figure 3-7 – Recovery of target analytes and additionally detected features in blank using different area ratio thresholds for blank subtraction..... - 50 -
- Figure 4-1 – Workflow of the data treatment, including LC-HRMS analysis (SCIEX OS), non-target data processing (MarkerView) and trend analysis (MATLAB). Before trend analysis, for target data, data fusion was performed. Subsequently, two complementary multivariate strategies of multivariate modelling were performed: 1. Column-wise PCA was performed on the data matrix having features in rows and time points in columns with Tscore<sub>r</sub> ranking of features. 2. Row-wise PCA and group-wise PCA (GPCA) were performed on the matrix of data where rows are time points and columns are features. Here, Tscore<sub>r</sub> ranked the time points and the groups of features were obtained. For real WWTP influent samples, the workflow of GPCA can further follow two possible routes, prioritising GPCs in decreasing variance order or prioritising increased trends due to environmental impacts. Additionally, for wastewater samples, a univariate statistic was applied as pre-filtering step. Trend analysis was followed by the identification of prioritised time trend relevant features (SCIEX OS, ChemSpider, MetFrag). ..... - 88 -
- Figure 4-2 – Score plot obtained by PCA of target data matrix of (a) principal component 2 (PC2 25.47%) versus principal component 1 (PC1 39.99%). The ellipses in the figure highlight groups of trends exemplarily shown in (b)..... - 92 -
- Figure 4-3 – Illustration of MEDA map for the set of target data, plotting feature index against each other. Colours in the plot reflect the level and direction (positive: red, or negative: blue) of the correlation between features. The red-blue box on the lower right part of the plot illustrates the features 110 to 149 with straight line trendless signals..... - 94 -
- Figure 4-4 – MEDA map fitted with 15 components for pre-filtered WWTP influent data matrix for positive ionisation mode. (a) and (b) show the sub-maps related to decreasing and increasing trend features. (c) Tscore<sub>r</sub> plot obtained by column-wise PCA modelling of pre-filtered WWTP influent data matrix in positive ionisation mode. .... - 97 -
- Figure 4-5 – Score vectors of six components from the GPCA model for the pre-filtered WWTP influent data matrix for positive ionisation mode. The prioritisation order is GPC6, GPC3, GPC1, GPC2, GPC5 and GPC4, respectively..... - 98 -
- Figure 4-6 – Identification of the lactam N-methylpyrrolidone (Feature 227, [M+H]<sup>+</sup>). The extracted ion chromatograms (XIC) of sample and standard are shown in (a). In (b) the N-methylpyrrolidone structure are presented assigned by MarvinSketch (v.19.27, ChemAxon) and referenced by Sigma-Aldrich. The estimated wastewater intensity pattern over time is illustrated in (c). Tandem mass spectrometry spectra used for identification of N-methylpyrrolidone with the corresponding standard spectrum shown in (d)..... - 101 -
- Figure 5-1 – Flowchart of the trend detection workflow applied for the assessment of the industrial WWTP based on spatial and temporal treatment behaviour..... - 138 -
- Figure 5-2 – Schematic representation of the multiple data collection for the used data matrix of all aligned features (F) of all sampling sites (SP) and measured time points (tp)... - 140 -
- Figure 5-3 – Mass-to-charge ratio (m/z) vs retention time scatter plots of aligned features by NTS over five months. In (a) the municipal wastewater (SP 1), in (b) the composite

- wastewater of the industrial area (SP 2), in (c) the composite industrial wastewater directly before the WWTP (SP 3), in (d) the wastewater within the WWTP (SP 4) and in (e) the WWTP effluent (SP 5) are shown. In (b) and (c) some homologous series are highlighted by red lines. .... - 144 -
- Figure 5-4 – Score plot of the PCA of the data matrix having sampling sites in the row, samples in the column and the intensity of features averaged over five months..... - 145 -
- Figure 5-5 – Rarity score of all detected features (8,931 in total, see Table 5-1) based on Krauss et al. (2019). The scores are sorted increasingly on a logarithmic scale..... - 147 -
- Figure 5-6 – Result of treatment efficiency of the WWTP. The comparison of the process WWTP effluent and influent of samples taken over five months is shown for the first sampling time, the average and for the last sampling time. Evaluation is based on Bader et al. (2017). The quotient of SP 5/SP 3 was calculated by dividing the intensities of, e.g. first time point by the intensity of, e.g. second time point. Details of every sampling time are shown in the Appendix (see 5.6.1.3). .... - 149 -
- Figure 5-7 – Result of treatment efficiency of the WWTP of studied five months. The number of aligned features in the different categories was based on fold change determined by pairwise comparisons of sampling sites (SP), respectively..... - 150 -
- Figure 5-8 – Comparison of influent and effluent WWTP samples over time, demonstrating the treatment efficiency of the industrial WWTP. The three parts of the figure correspond to features prioritised for WWTP influent samples (SP 3) according to a decreasing trend in intensity over time which also withstand treatment processes (still detectable in SP 5). The given mass error refers to the accurate mass of aligned features of influent and effluent samples. The corresponding colour scheme illustrates the pairwise comparison of the process influent and effluent (SP 3 to SP 5) shown in influent data, resulting in categorisation of features into elimination, decrease or consistency based on the calculated fold change (Bader et al., 2017). For completeness, the features detection in sampling site SP 4 is shown, too. .... - 152 -
- Figure 5-9 – MS/MS-spectra of three detected peaks after treatment process out of temporal prioritised features of WWTP influent samples. Retention times (RT), accurate mass (m/z), as well as proposed sum formulae (based on 'Non-targeted Screening' workflow of SCIEX OS and ChemSpider data) with a relative mass deviation of the exact mass from the proposed sum formulae, are shown. .... - 154 -
- Figure 6-1 – Flowchart of the industrial wastewater sample analysed by the offline 2D LC system. For structure confirmation, the wastewater sample was measured again for enrichment using LC-HRMS and analysed by <sup>1</sup>H-NMR..... - 174 -
- Figure 6-2 – LC-HRMS results in ESI (-) of the feature with ID 4; (a) Signal intensity comparison of feature 4 (XIC m/z 250.9265) for eluent fraction of wastewater and blank sample; (b) MS spectrum of precursor-ion of m/z 250.9265 with isotopic pattern for di-chlorination and (c) MS/MS spectrum of precursor-ion of m/z 250.9265 of eluent fraction. .... - 178 -
- Figure 6-3 – Proposed structures for formula finder result C<sub>6</sub>H<sub>2</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>5</sub> for the feature with ID 4: (a) 2,5-dichloro-4,6-dinitro-phenol, (b) 3,4-dichloro-2,6-dinitro-phenol and (c) 2,3-dichloro-4,6-dinitro-phenol..... - 179 -
- Figure 6-4 – Comparison of authentic standard 3,4-dichloro-2,6-dinitrophenol (upper row, 50 µg L<sup>-1</sup>) and eluent fraction of wastewater sample (lower row); UV spectra of (a) authentic standard (Ref.) and (d) of eluent fraction (EF); cut-out of MS spectrum of (b) precursor-ion m/z 250.9265 of authentic standard and (e) of precursor-ion m/z 250.9265 of eluent fraction

- showing identical isotopic pattern; MS/MS spectrum of (c) precursor-ion m/z 250.9265 of authentic standard and (f) of precursor-ion m/z 250.9265 of eluent fraction. .... - 180 -
- Figure 6-5 – Comparison of <sup>1</sup>H-NMR spectra of (a) authentic standard 3,4-dichloro-2,6-dinitrophenol, (b) eluent fraction of wastewater sample and of (c) blank. The chemical shift of around 8 ppm belongs to those of aromatic compounds. .... - 181 -
- Figure-A 3-1 – Illustration of different fragmentation methods (DDA  $\triangleq$  data-dependent acquisition, vDIA  $\triangleq$  variable data-independent acquisition) used in the method development. .... - 61 -
- Figure-A 3-2 – Illustration of stability of ISTD (exemplarily shown for Diuron-D6) over one batch for quality control. .... - 62 -
- Figure-A 3-3 – Comparison of the CCF from a) an exemplary undiluted pre-treated wastewater sample (SP 4, first chromatogram) and the same sample with dilution (1 to 10 with Eluent A/B (V/V), second chromatogram) and b) an exemplary WWTP influent sample (SP 2) without any dilution (first chromatogram) and diluted 1 to 10 (second chromatogram). The y-axis presents the ITC calculated by the CCF divided by the accumulation time (in the given case: 0.1 s)..... - 65 -
- Figure-A 3-4 – Recovery of the WWTP inlet (SP 3) and the WWTP outflow (SP 5) of authentic samples from January 2018 (2018-01-01 to 2018-01-31) after six months and 15 months. The recovery is based on all features (target approach) which were extracted in January 2018 in the suspected-target evaluation. The data label of bars indicates the defined sample amount covering the recovery in % (cf. y-axis). The bars illustrate the recovery in %. The results represent the mean values of 31 samples for each matrix. .... - 67 -
- Figure-A 4-1 – Scatter plot of Q-values vs Hotelling's T2 values obtained by PCA analysis of column-wise validation (target) data..... - 119 -
- Figure-A 4-2 – Tscore plot obtained by row-wise PCA processing of validation data.... - 121 -
- Figure-A 4-3 – Score vectors of first 13 components from the GPCA model for the validation (target) data set (a). Scores and loadings for the six prioritised decreasing/increasing trends using groups obtained from MEDA (b). .... - 122 -
- Figure-A 4-4 – Total number of feature groups and mean number of features per group as a function of selected threshold  $\gamma$  from the correlation map. The smaller the mean size of the group, the greater the sparsity. Nevertheless, by increasing the threshold value, some trends will be missed. So, initial experience about the association patterns of mass fragments is significant to select threshold value. .... - 123 -
- Figure-A 4-5 – Example of MK failed monotonic time trend test. Patterns with fluctuation in increasing/decreasing part of the trend cause failure in the MK test. The example shows the feature with ID 133 in positive ionisation mode. The p-value is 0.069, indicating no trend, according to MK ( $p > 0.05$ )..... - 126 -
- Figure-A 4-6 – 2D Score plot of the first two components derived from a PCA model based on LC-HRMS data matrix of influent WWTP sample in positive ionisation mode. Column-wise PCA model of pre-processed initial data matrix of positive ionisation mode, used 25 components to explain 80% of the total variance in the data..... - 126 -

- Figure-A 4-7 – 2D score (a & b) and loading (c & d) plots obtained by PCA modelling of pre-filtered WWTP influent data matrices containing increasing/decreasing intensity patterns for positive and negative ionisation mode, respectively. .... - 127 -
- Figure-A 4-8 – Tscore values for PCA modelling of (a) initial WWTP influent data matrix (69\*2380) in positive ionisation mode using 15 PCs with a total variance of 81.1% and (b) prefiltered data matrix (69\*234) based on 10 PCs with a total variance of 87.3%. The most relevant time points regarding high score values are indicated by dashed circles. .... - 128 -
- Figure-A 4-9 – 2D score plot of first 2 PCs in GPCA of pre-filtered WWTP influent data in positive ionisation mode (a) and the corresponding loading (line/bar) plot (b). .... - 129 -
- Figure-A 4-10 – Results of GPCA of positive ionisation mode. For example, the first score vector with the highest variance has a decreasing trend. The second GPC contains the most important features responsible for increasing trends, especially at time point/sample s48. . .... - 129 -
- Figure-A 4-11 – Results of GPCA of negative ionisation mode. All GPCs belonged to decreasing trends. .... - 130 -
- Figure-A 5-1 – Result of quality control during data analysis using the evaluation scheme of Bader et al. (2017). The comparison of WWTP effluent and influent of samples taken over five months is shown focussing on spiked ISTD. Evaluation is based on Bader et al. (2017). The quotient of SP 5/SP 3 was built by dividing every day's intensity from another... - 161 -
- Figure-A 5-2 – Result of treatment efficiency of the WWTP. The pairwise comparison of process influent and effluent samples taken over five months, resulted in features assigned to the categories elimination, decrease, consistency, increase, or formation samples. The categorisation was calculated by the fold change (fc) approach introduced by Bader et al. (2017). .... - 165 -
- Figure-A 5-3 – Result of treatment efficiency of the WWTP considering the flow time in the WWTP. The assessment scheme of pairwise comparison introduced by Bader et al. (2017) was modified by calculating the fold change (fc) between the intensities of process influent sample of the studied day and the intensities of process effluent of the next day to check whether some features had a delay in flow time in the WWTP. .... - 166 -
- Figure-A 6-1 – Instrument Resolution of SCIEX qTOF x500R (Loos, 2018). .... - 185 -
- Figure-A 6-2 – Acid-based gradient for development of HPTLC plates. .... - 189 -
- Figure-A 6-3 – Alkaline-based gradient for development of HPTLC plates. .... - 189 -
- Figure-A 6-4 - Ether-based gradient for development of HPTLC plates. .... - 190 -
- Figure-A 6-5 – Representation of ideal signal cutting (dashed lines) at full width at half maximum (FWHM) of normal distributed signal peak. .... - 191 -
- Figure-A 6-6 – Constitutional isomers of structural proposal dichlorodinitrophenol for unknown. .... - 192 -
- Figure-A 6-7 – Densitogram of reference material 3,4-dichloro-2,6-dinitrophenol (50 ng) measured using acid-based gradient under UV. The colours show the respective wavelengths at which measurements were taken in parallel. .... - 193 -

- Figure-A 6-8 – Densitogram of unknown substance measured using acid-based gradient under UV. The colours show the respective wavelengths at which measurements were taken in parallel. .... - 194 -
- Figure-A 6-9 – Assessment of substance toxicity by HPTLC. In a) the densitogram (recorded at 240 nm) and in b) the toxicity result of eluent fraction of wastewater sample are recorded using 'alkaline' gradient (see App. 1c, Figure-A 6-2 to Figure-A 6-4). The black box highlights the black spots indicating bioluminescence quenching and consequently toxicity of peak of interest. The placement of TLC plate is listed in App. 1d, Table-A 6-2..... - 195 -
- Figure-A 6-10 – Calibration curve of lower concentration range and higher concentration range in mobile phase A of LC-HRMS (Milli-Q® with 0.1% formic acid) of reference material 3,4-dichloro-2,6-dinitrophenol for quantification of isolated material using LC-UV. .... - 196 -
- Figure-A 6-11 – Determination of average quantified (by authentic standard 3,4-dichloro-2,6-dinitrophenol in Milli-Q® water) concentration of identified 'known unknown' in wastewater samples of industrial plant and the WWTP in mg L<sup>-1</sup>. .... - 196 -

### 8.3 List of Tables and Tables-A

- Table 1-1 – Adducts and fragments typically occurring in ESI ionisation (Keller et al., 2008)...  
..... - 17 -
- Table 3-1 – List of tested columns with their specification and properties (TMS  $\triangleq$  the functional group trimethylsilyl)..... - 39 -
- Table 3-2 – Results of column comparison based on suspected-target analysis of a triplicate injection of 10  $\mu\text{g L}^{-1}$  standard for the four different columns. The following parameters were compared: retention time factor  $k$ , asymmetry factor  $AF$ , the full width at half maximum (FWHM) and standard deviation (SD) of the relative retention time (RRT). The same method was used for all four columns. A high  $k$ -value indicates that the analyte is highly retained and has spent a significant amount of time interacting with the stationary phase,  $AF$  near 1 indicates symmetric peak ( $AF > 1 \triangleq$  tailing,  $AF < 1 \triangleq$  fronting), small FWHM represents sharp peaks and small standard deviations of RRT shows stable chromatographic separation conditions. .... - 45 -
- Table 3-3 – Result of unknown screening of authentic wastewater samples. True positives/negatives (TP / TN), false positives (FP) and false negatives (FN) were assigned to get an impression of the suitability of the non-target method. The quality parameters are given in % and are assigned based on suspected-target list matches of treated (SP 5), pre-treated (SP 4) and untreated (SP 2, SP 3) wastewater samples. Overall, 174 samples were measured in positive ionisation mode and 108 were measured in negative ionisation mode.  
..... - 47 -
- Table 4-1 – GPCA results for feature prioritisation of WWTP influent samples containing the whole subset of continuously increasing trends of positive ionisation mode. .... - 99 -
- Table 5-1 – Table of different sampling sites (SP) across the treatment processes of the WWTP during November 2018 to March 2019 used for spatial-temporal trend analysis. The corresponding schematic overview is shown in chapter 3 of the dissertation. .... - 138 -

---

Table 5-2 – Categories applied for the assessment of wastewater treatment processes. The calculated fold chain (fc) is based on feature's intensity ( $fc = \text{intensity}_{\text{Effluent}} / \text{intensity}_{\text{Influent}}$ ) across treatment process. Categories based on Bader et al. (2017). .....	- 141 -
Table 5-3 – Aligned features initially detected within five months and after applying componentisation by GPCA introduced in chapter 4. ....	- 143 -
Table 6-1 – Data evaluation workflow of detected features filtration for positive (ESI (+)) and negative ionisation (ESI (-)) mode; RT $\triangleq$ retention time, auto $\triangleq$ software-based filter, manual $\triangleq$ manual exclusion of features. ....	- 177 -
Table 6-2 – Remaining features in negative ionisation mode (ESI (-)) of non-target LC-HRMS analysis after applying blank subtraction, exclusion of features without MS <sup>2</sup> data and false positives, adduct grouping and retention time filtering. ....	- 178 -
Table 6-3 – Excerpt of empirical formulae of the feature with ID 4 assigned by Formula Finder (SCIEX OS) in consideration of operation plant information with maximal element options of C <sub>0-30</sub> , H <sub>0-30</sub> , O <sub>0-5</sub> , N <sub>0-5</sub> , Cl <sub>0-3</sub> , S <sub>0-3</sub> and a mass tolerance of 5 ppm. ....	- 179 -
Table-A 3-1 – Table of analytes and ISTD used for method development, method validation and quality control (QC). All data are rounded data. Fragment ions are listed in decreasing intensity from 1 to 5. RT $\triangleq$ Retention time (*industrial chemical). ....	- 57 -
Table-A 3-2 – Table of concentrations in $\mu\text{g L}^{-1}$ of spiked compounds for the development of the data evaluation method. ....	- 60 -
Table-A 3-3 – Parameter settings of 'non-target analysis' determination with SCEIX OS. ....	- 60 -
Table-A 3-4 – Table of settings used for data-independent acquisition (vDIA). ....	- 61 -
Table-A 3-5 – Method validation results of internal standards in positive (+) and negative (-) ionisation mode measured over seven days. ....	- 63 -
Table-A 3-6 – Results of the experiments of centrifugation of 17 inflow samples of the WWTP (SP 3), 12 pre-treated samples (SP 4) and 16 samples of outflow of the WWTP (SP 5). The relative standard deviation (RSD) is the mean of nine measurements and represents the deviation of centrifugation experiments to initial measurements without centrifugation. ....	- 63 -
Table-A 3-7 – Results of stability analysis during sample storage. The storage conditions of freezing and cooling were tests for samples of sampling sites 3 to 5. In the table, the recovery in % is presented. ....	- 66 -
Table-A 3-8 – Tested mobile phases during method development. ....	- 67 -
Table-A 3-9 – Recovery of target analytes during optimisation of LC-column. In the table, the mean areas of three injections of $10 \mu\text{g L}^{-1}$ standard mixture with corresponding standard deviation (SD) are shown. The results based on mobile phase 2 (see Table-A 3-8). ....	- 68 -
Table-A 3-10 – Retention (capacity) factor ( $k = (tR - t0)/t0$ ) of target analytes during optimisation of LC-column. In the table, the mean k of three injections of $10 \mu\text{g L}^{-1}$ standard with corresponding standard deviation (SD) are shown. The void time for calculation based on each column on the injection peak (0.9 min). The capacity factor shows the retention of an analyte on the chromatographic column. A high k-value indicates that the analyte is	



- highly retained and has spent a significant amount of time interacting with the stationary phase. The results based on mobile phase 2 (see Table-A 3-8)..... - 69 -
- Table-A 3-11 – Asymmetric factor (AF) of target analytes during optimisation of LC-column. In the table, the mean AF of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) is shown. The asymmetry factor shows the distance from the centre line of the peak to the back slope, divided by the distance from the centre line of the peak to the front slope, with all of the measurements made at 10% of the maximum peak height ( $\text{AF} = 1 \triangleq$  symmetric peak,  $\text{AF} > 1 \triangleq$  tailing,  $\text{AF} < 1 \triangleq$  fronting). The results based on mobile phase 2 (see Table-A 3-8). ..... - 70 -
- Table-A 3-12 – Results of full width at half maximum (FWHM) of target analytes during optimisation of LC-column. In the table, the mean FWHM of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) is shown. The FWHM shows the chromatographic peak width, in minutes, of the detected peak measured at half of its apex intensity. The results based on mobile phase 2 (see Table-A 3-8). ..... - 71 -
- Table-A 3-13 – Results of relative retention time (RRT) target analytes during optimisation of LC-column. In the table, the mean relative RT of three injections of  $10 \mu\text{g L}^{-1}$  standard with corresponding standard deviation (SD) is shown. The RRT shows the ratio of the target analyte's retention time to the retention time of ISTD (diuron-D6 in positive and mecoprop-D3 in negative ionisation mode). The results based on mobile phase 2 (see Table-A 3-8).. ..... - 72 -
- Table-A 3-14 – Recovery of target analytes in industrial wastewater (SP 3) using DDA and DIA for LC-HR-MS/MS. The comparison was made using three replicates (Rep. 1 to Rep 3). 'N.D.'  $\triangleq$  the target analyte was not detected. .... - 73 -
- Table-A 3-15 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the intensity threshold of 100 cps. 'Identified' means the target was detected in all samples within the sampling site. .... - 74 -
- Table-A 3-16 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the intensity threshold of 500 cps. 'Identified' means the target was detected in all samples within the sampling site. 'N.D.' means the target was not detected in the industrial wastewater samples. .... - 75 -
- Table-A 3-17 – Results of recovery of target analytes for spiked industrial wastewaters (see Figure 3-1) and ultra-pure water (UPW) for the threshold of 1,000 cps. 'Identified' means the target was detected in all samples within the sampling site. 'N.D.' means the target was not detected in the samples. .... - 76 -
- Table-A 3-18 – Results of Retention time and mass error tolerance.  $\text{RT} \triangleq$  Retention time. .... - 77 -
- Table-A 3-19 – Result of method validation of suspected-targets of 100 un-spiked wastewater sample (SP 5) in positive (+) and negative (-) ionisation mode. Previously, the samples were analysed with LC-UV declaring them positive for at least one suspected-target analyte..... - 78 -
- Table-A 3-20 – Exemplary feature table of treated wastewater (SP 5) in positive ionisation mode (+). ..... - 79 -
- Table-A 4-1 – Table of data sets for data analysis method development and validation. - 108 -

---

Table-A 4-2 – Spiked target analytes for training data set covering relevant (black) and non-relevant (grey) time trends (n= 29). All data are rounded data and refer to the LC-HRMS measurement. For the identification, a maximum mass error of $\pm 2$ ppm and a maximum retention time error of $\pm 0.05$ min were approved (*industrial chemical). .....	- 109 -
Table-A 4-3 – Table of used internal standards (ISTD). All data are rounded data and refer to the LC-HRMS measurement. For the identification, a maximum mass error of $\pm 2$ ppm and a maximum retention time error of $\pm 0.05$ min were approved. Fragment ions are listed in decreasing intensity from 1 to 5. Pos $\triangleq$ positive ionisation mode, neg $\triangleq$ negative ionisation mode. RT $\triangleq$ Retention time. ....	- 112 -
Table-A 4-4 – LC-HRMS acquisition method. ....	- 113 -
Table-A 4-5 – Table of analytes and ISTD used for quality control (QC). All data are rounded data. Fragment ions are listed in decreasing intensity from 1 to 5. RT $\triangleq$ Retention time, R $\triangleq$ Recovery. ....	- 114 -
Table-A 4-6 – List of labelled internal standards screened in positive and negative ionisation mode, detected ion/adduct, retention time, sample matches (n= 29) and trend. ....	- 118 -
Table-A 4-7 – Detailed result of the 30 prioritised features (feature index, m/z and retention time) of column-wise PCA (the result of 13PCs) in combination with Tscore <sub>f</sub> index. 'Both trends' means first decreasing in intensity followed by increasing intensity changes over time. The spiked target analytes recovered are listed. Features assigned as unknowns did not fit spiked target compounds. ....	- 120 -
Table-A 4-8 – Groups of recorded features by GPCA modelling of validation (target) data. ....	- 124 -
Table-A 4-9 – Detailed list of features prioritised by GPCA. The features printed in thick are these ones following the relevant trend of continuously increasing/decreasing intensity. In total, 130 relevant features were prioritised. ....	- 130 -
Table-A 4-10 – List of m/z, retention time and the trend of prioritised features using GPCA of positive and negative ionisation mode (cf. Table S-9). ....	- 131 -
Table-A 5-1 – Targets and ISTD used for testing data set covering relevant and non-relevant time trends (*industrial chemical). ....	- 158 -
Table-A 5-2 – LC-HRMS acquisition method. ....	- 159 -
Table-A 5-3 – MarkerView was used for peak detection, peak alignment, normalisation of retention times and blank value subtraction. The following criteria were used. The average values of peak intensities and retention times of internal standards (ISTD, see 5.6.1.1) were used to correct the retention times (RT) and to normalise the peak intensities. ....	- 160 -
Table-A 5-4 - Site-specific features of sampling stations SP 1 to SP 5 The mean intensity was averaged over all sampling times. RT $\triangleq$ retention time. ....	- 162 -
Table-A 5-5 – Prioritised features by rarity score introduced by Krauss et al. in 2019. RT $\triangleq$ retention time. ....	- 164 -
Table-A 5-6 – Features which were identified in both the WWTP influent (SP 3) and treated wastewater (SP 5) which consequently withstand treatment process. The m/z, retention time (RT), median intensity over all measured wastewater samples, as well as the corresponding trend are listed. ....	- 167 -

Table-A 5-7 – Features which withstand the treatment process compared to the WWTP influent (SP 3) and their corresponding trends. In (a) the features of the WWTP influent are shown which have an increasing trend, in (b) these features of the WWTP influent which have a decreasing trend and in (c) the corresponding features of the WWTP effluent with their trends are listed describing these which withstand treatment process. .... - 170 -

Table-A 6-1 – Table of analytes and ISTD used for method development. Additionally, the data of quality control (QC) is listed. All data are rounded data and refer to the LC-HRMS measurement (2<sup>nd</sup> dimension). For the identification, a maximum mass error of  $\pm 2$  ppm and a maximum retention time error of  $\pm 0.05$  min were approved. Fragment ions are listed in decreasing intensity from 1 to 5. RT  $\triangleq$  Retention time, Rec.  $\triangleq$  Recovery (\* industrial chemical). .... - 186 -

Table-A 6-2 – Description of schematic order for toxicity testing (Used bacteria: *Aliivibrio fischeri*). .... - 190 -

## 8.4 List of Publications

### 8.4.1 Articles in Peer-Reviewed Journals

Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Sci. Total Environ.* 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>

Purschke, K., Vosough, M., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Evaluation of non-target long-term LC-HRMS time series data using multivariate statistical approaches. Submitted to *Anal. Chem.*

### 8.4.2 Other Articles

Purschke, K., Weber, M., Schmidt, T.C., 2019. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. *Vom Wasser*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. 117 (2), 33–72, ISSN 0083-6915

Purschke, K., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Mit modernsten Analysenmethoden auf Spurensuche. *LABO.* 52 (4), 16–19.

### 8.4.3 Oral Presentations

Purschke, K., Weber, M., Schmidt, T.C., 2019. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. Short presentation, *Wasser 2019, Jahrestagung der Wasserchemischen Gesellschaft*; Erfurt, Germany, 27/05–29/05/2019.

Purschke, K., Weber, M., Schmidt, T.C., 2020. Offline 2D chromatography and non-target screening to identify unknowns in industrial wastewater. 30. Doktorandenseminar des AK Separation Science. Hohenroda, Germany, 11/01–14/01/2020.

### 8.4.4 Poster presentations

Purschke, K., Weber, M., Schmidt, T.C., 2018. Validation of a suspect/non-target screening method for industrial wastewater. 3. Mülheimer Wasseranalytisches Seminar; Mülheim an der Ruhr, Germany, 12/09–13/09/2018

Purschke, K., Weber, M., Schmidt, T.C., 2019. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. ANAKON 2019; Münster, Germany, 25/03–28/03/2019

Purschke, K., Weber, M., Schmidt, T.C., 2019. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. Wasser 2019, Jahrestagung der Wasserchemischen Gesellschaft; Erfurt, Germany, 27/05–29/05/2019

Purschke, K., Weber, M., Schmidt, T.C., 2019. LC-HRMS data handling strategy for non-target monitoring approaches of industrial wastewater. Langenauer Wasserforum, Langenau, Germany, 12/11–13/11/2019

## 8.5 Declaration of Scientific Contribution

The present thesis includes work that has been published in cooperation with co-authors, with my own contributions declared as follows:

### ***Chapter 4***

Purschke, K., Vosough, M., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Evaluation of non-target long-term LC-HRMS time series data using multivariate statistical approaches. Submitted to Anal. Chem.

Experiments were planned, implemented and controlled by Kirsten Purschke. Data evaluation were carried out predominantly by Kirsten Purschke and partially by Maryam Vosough. The Draft manuscript was written by Kirsten Purschke. Juri Leonhardt and Torsten C. Schmidt were responsible for the editorial of the manuscript. Torsten C. Schmidt and Markus Weber supervised this study.

### ***Chapter 6***

Purschke, K., Zoell, C., Leonhardt, J., Weber, M., Schmidt, T.C., 2020. Identification of unknowns in industrial wastewater using offline 2D chromatography and non-target screening. Sci. Total Environ. 706. <https://doi.org/10.1016/j.scitotenv.2019.135835>

Kirsten Purschke planned, implemented, evaluated and controlled the research workflow. Christian Zoell planned, implemented and evaluated the research part of toxicological determination using HPTLC. Juri Leonhardt and Torsten C. Schmidt were responsible for the editorial of the manuscript. Torsten C. Schmidt and Markus Weber supervised this study.

## **8.6 Curriculum Vitae**

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

## 8.7 Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

***„Environmental Forensics of Industrial Wastewater  
based on Non-Target Screening“***

selbst verfasst, keine außer den angegebenen Hilfsmitteln und Quellen benutzt habe, alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet sind und die Arbeit in dieser oder ähnlicher Form noch bei keiner anderen Universität eingereicht wurde.

Leverkusen, im Juni 2020



## 8.8 Danksagung

Mein größter Dank gilt Prof. Dr. Torsten C. Schmidt, der seitens der Universität Duisburg-Essen die Anfertigung meiner Thesis ermöglicht und begleitet hat. Unsere hilfreichen Diskussionen über Ergebnisse und seine wertvollen Rückmeldungen zur Verbesserung von Forschungsartikeln waren eine unverzichtbare Unterstützung auf dem Weg zu der vorgelegten Dissertation.

Für die Möglichkeit dieser Arbeit, das gespendete Vertrauen, die Anleitung und Betreuung, die zu dem Gelingen meiner These beigetragen haben, bedanke ich mich ganz herzlich bei Dr. Markus Weber und Dr. Hans-Christian Mans.

Prof. Dr. Oliver J. Schmitz danke ich für die Übernahme des Zweitgutachtens.

Meinen Koautoren der veröffentlichten Forschungsartikel Prof. Dr. Maryam Vosough und Christian Zoell möchte ich für die hervorragende Zusammenarbeit und den fachlichen Austausch meinen Dank aussprechen. Christian Zoell danke ich weiter für die Toxizitätstests und Prof. Dr. Maryam Vosough für die kompetente Unterstützung bei der Programmierung in MATLAB®.

Dr. Thomas Westfeld danke ich für die Durchführung der NMR-Messungen und Unterstützung bei der Auswertung der Ergebnisse.

Mein besonderer Dank gilt Dr. Juri Leonhardt, Lotta L. Hohrenk, Wiebke Kaziur-Cegla, Sarah Neumann und Janine Piela für die umfangreiche Durchsicht meiner Dissertation.

Darüber hinaus möchte ich mich bei allen Freunden und Kollegen des Überwachungslabors der Umweltanalytik für die tolle Unterstützung und die äußerst angenehme Arbeitsatmosphäre bedanken.

Herzlich danke ich Jens und meiner Familie für ihre nicht nachlassende Ermutigung während der Doktorarbeit und des gesamten Studiums.