Mining of LC-HRMS data for the assessment
of water treatment processes

Kumulative Dissertationsschrift

zur Erlangung des akademischen Grades doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt der Fakultät Nachhaltigkeit

der Leuphana Universität Lüneburg

Tobias Bader

geboren am 14. August 1989 in Weißenhorn

2018

PREFACE

The underlying work for this cumulative dissertation was conducted in the Laboratory for Operation Control and Research, Zweckverband Landeswasserversorgung in cooperation with the Institute of Sustainable and Environmental Chemistry (Prof. Dr. Kümmerer), Leuphana University of Lüneburg between October 2014 and October 2017. The work was funded by Zweckverband Landeswasserversorgung.

This cumulative dissertation is based on three scientific contributions. Two articles are published in international peer-reviewed journals. The third contribution, a peer-reviewed book chapter, is published in the ACS Symposium Series.

- o Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2016**). General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta* 935: 173-186.

    DOI: 10.1016/j.aca.2016.06.030

    [Hereinafter be referred to as 'Paper 1']

- o Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2017**). LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes. *Analytical Chemistry* 89: 13219-13226.

    DOI: 10.1021/acs.analchem.7b03037

    [Hereinafter be referred to as 'Paper 2']

- o Tobias Bader, Wolfgang Schulz, Thomas Lucke, Wolfram Seitz, Rudi Winzenbacher (**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* (Vol. 1242, pp. 49-70): American Chemical Society.

    DOI: 10.1021/bk-2016-1242.ch003

    [Hereinafter be referred to as 'Book Chapter']

All articles are reprinted with kind permission of Elsevier and the American Chemical Society at the end of this thesis (see 'Reprint of articles included in this thesis').

In the following extended summary, an introduction and the objectives of this work are described in chapters 1 and 2. The developed methods are stated in chapter 3 and the main results of the studies are presented and discussed in chapter 4, followed by the final conclusion and an outlook to future research purposes in chapter 5.

ABSTRACT

The emission of anthropogenic trace substances into the aquatic environment continuously poses challenges to water suppliers. The contamination of raw waters with organic trace substances requires complex water treatment processes to secure drinking water quality. The routine monitoring of these raw waters as well as the behavior and fate of organic trace substances during different treatment processes is of great interest to recognize and counter potential dangers at an early stage. Chromatographic separation techniques coupled to triple quadrupole mass spectrometers are conventionally used for the reliable monitoring of traces of known polar substances. However, such analytical techniques usually fail to recognize unknown compounds. This weakness presents a serious restriction with regard to the monitoring of treatment processes, since transformation products are often not - or not sufficiently - characterized and are thus only detected sporadically. Non-target screening using liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) allows the detection of thousands of compounds within a single run and covers known as well as unknown substances. Compared to the established analytical techniques, this is a decisive advantage for the monitoring of raw and process waters during water treatment. While the analytical technique LC-HRMS has undergone significant developments in recent years, the algorithms for data processing reveal clear weaknesses.

This dissertation therefore deals with reliable processing strategies for LC-HRMS data. The first part of this work seeks to highlight the problematics of false positive and false negative findings. Based on repeated measurements, various strategies of data processing were assessed with regard to the repeatability of the results. To ensure that real peaks were barely or not removed by the filtering procedure, samples were spiked with isotope-labeled standards. The results emphasize that the processing of sample triplicates results in sufficient repeatability and that the signal fluctuation across the triplicates emerged as a powerful filtering criteria. The number of false positives and false negatives could be significantly reduced by the developed strategies which consequently improve the validity of the data.

The second part of this thesis addresses the development of processing strategies particularly aimed at assessing water treatment processes. The detected signals were tracked across the treatment process and classified based on their fold changes. A more reliable signal classification was achieved by implementing a recursive integration approach. Special integration algorithms allow a reliable signal classification even though the signal to be compared was below the intensity threshold. Different combinations of replicates of process influents and effluents were processed for evaluating the repeatability. The good repeatability was indicated by the results of both the plausibility checks and the ozonation process (ozonation of pretreated river water) and thus points to high reliability. The

developed procedure enables the assessment of water treatment processes based on the changes in the pattern of all detected signals and offers a more comprehensive picture of the treatment efficiency. Particularly with regard to transformation products, existing knowledge gaps can be reduced by this approach, albeit the entire variety of chemicals cannot be covered completely.

The applicability of the developed strategies to real world applications is demonstrated in the last part of this work. Besides the prioritization of the generated results, the main focus was the identification of recognized compounds.

The developed strategies clearly improve the validity of the underlying data. The combination of LC-HRMS analysis with reliable processing strategies opens up multiple possibilities for a more comprehensive monitoring of water resources and for the assessment of water treatment processes. The processing strategies and validation concepts may be easily transferred to other research fields.

ZUSAMMENFASSUNG

Die Emission von anthropogenen Spurenstoffen in die aquatische Umwelt stellt Wasserversorger fortwährend vor neue Herausforderungen. Mit organischen Spurenstoffen belastete Rohwässer erfordern komplexe Aufbereitungsverfahren zur Sicherung der Trinkwasserqualität. Sowohl die routinemäßige Überwachung dieser Rohwässer als auch das Verhalten organischer Spurenstoffe während verschiedener Aufbereitungsprozesse ist von großem Interesse, um etwaige Gefahren frühzeitig zu erkennen. Herkömmlich verwendete Analysetechniken - meist chromatographische Trenntechniken gekoppelt an Triple-Quadrupol-Massenspektrometer - erlauben die Überwachung bekannter polarer Substanzen im Spurenbereich, versagen jedoch meist beim Erkennen unbekannter Verbindungen. In Hinblick auf die Überwachung von Aufbereitungsprozessen stellt diese Schwachstelle eine große Einschränkung dar, da Transformationsprodukte oftmals nicht oder nur unzureichend charakterisiert bzw. bekannt sind und demnach nur sporadisch erfasst werden können. Mit sogenannten „*non-target screening*"-Methoden lassen sich tausende Verbindungen in einer Analyse erfassen. Durch die Kopplung von Flüssigkeitschromatographie mit hochauflösender Massenspektrometrie (LC-HRMS) können neben bekannten auch unbekannte Spurenstoffe detektiert werden. In Hinblick auf die Überwachung von Roh- und Prozesswässern bei der Wasseraufbereitung stellt dies einen entscheidenden Vorteil gegenüber etablierten analytischen Techniken dar. Während die LC-HRMS-Analytik in den letzten Jahren stark weiterentwickelt wurde, zeigen die Auswertealgorithmen jedoch noch deutliche Schwachstellen.

Die vorliegende Doktorarbeit befasst sich daher mit Strategien zur verlässlichen Auswertung von LC-HRMS-Daten. Im ersten Schritt wurde die Problematik von falsch-positiv- und falsch-negativ-Befunden beleuchtet. Basierend auf Wiederholmessungen wurden verschiedene Strategien der Datenauswertung hinsichtlich der Ergebniswiederholbarkeit bewertet. Durch das Dotieren isotopenmarkierter Standards wurde sichergestellt, dass echte Signale durch die Datenfilterung nicht entfernt werden. Die Daten zeigen, dass Probentriplikate hinreichend gute Ergebnisse liefern und die Berücksichtigung der Signalschwankungen über Replikatmessungen eine sehr effiziente Filtermethode darstellt. Durch die entwickelten Vorgehensweisen konnte die Anzahl von falsch-positiv- und falsch-negativ-Befunden bei der Datenauswertung signifikant reduziert und die Validität der Daten gesteigert werden.

Im zweiten Teil der Arbeit wurden Auswertestrategien speziell zur Bewertung von Wasseraufbereitungsprozessen entwickelt. Die detektierten Signale wurden während des Behandlungsprozesses verfolgt und einer Signalklassifikation unterzogen. Durch die Implementierung eines rekursiven Ansatzes unter Verwendung spezieller Integrationsalgorithmen konnte eine deutlich verlässlichere Signalklassifikation vorgenommen werden, auch wenn zu vergleichende Signale

unterhalb des festgelegten Intensitätsschwellenwertes lagen. Zur Bewertung der Wiederholbarkeit wurden verschiedene Kombinationen von Replikaten aus Prozesszu- und -abläufen ausgewertet. Sowohl die Ergebnisse der Plausibilitätskontrollen als auch die des realen Prozesses einer Ozonung von vorgereinigtem Flusswasser zeigten eine gute Wiederholbarkeit und lassen demnach auf eine hohe Reliabilität schließen. Diese Arbeit ermöglicht es, die Bewertung von Aufbereitungsprozessen basierend auf der Veränderung aller detektierbarer Signale vorzunehmen und dadurch ein deutlich umfangreicheres Bild der Prozesseffizienz zu erhalten. Besonders in Hinblick auf Transformationsprodukte können durch diese Herangehensweise bestehende Wissenslücken reduziert werden, wenngleich auch hierdurch die Vielfalt der chemischen Stoffe nicht vollständig abgedeckt werden kann.

Im letzten Teil der Arbeit wurde die Anwendbarkeit der entwickelten Strategien in der Realität verdeutlicht. Neben der Priorisierung von Ergebnissen wurde ein Hauptaugenmerk auf die Identifikation erfasster Komponenten gelegt.

Die in dieser Arbeit erarbeiteten Strategien tragen deutlich zur Steigerung der Datenvalidität bei. Die Kombination der LC-HRMS-Analytik mit verlässlichen Auswertestrategien eröffnet eine Vielzahl an Möglichkeiten zur umfassenderen Überwachung von Wasserressourcen und zur Beschreibung von Aufbereitungsprozessen. Die entwickelten Strategien und Validierungskonzepte lassen sich auf andere Forschungsgebiete übertragen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

C	Consistency (category for signal classification)
D	Decrease (category for signal classification)
DA	Discriminant analysis
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
E	Elimination (category for signal classification)
EIC	Extracted-ion chromatogram
ESI	Electrospray ionization (+/- for positive / negative mode)
F	Formation (category for signal classification)
fc	Fold change
GAC	Granular activated carbon
I	Increase (category for signal classification)
$\bar{I}$	Mean improvement factor
IS	Internal standard (here: isotope-labeled)
LC-HRMS	Liquid chromatography coupled to high-resolution mass spectrometry
LW	Zweckverband Landeswasserversorgung
MS	Mass spectrometry
MS²	Tandem mass spectrometry (i.e. MS/MS)
PCA	Principal component analysis
$\overline{RR}$	Mean rate of recognition
RSD	Relative standard deviation
SD	Standard deviation
SWATH	Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra
TPPO	Triphenylphosphine oxide
WWTP	Wastewater treatment plant

1 INTRODUCTION

A variety of tens of thousands of organic substances are in daily use and manufactured, consumed and disposed of in households or industry/agriculture. Various substance classes such as pharmaceuticals, pesticides, personal care products and industrial chemicals, among others, have the potential to enter the aquatic environment as micropollutants.[1-5] The presence of these micropollutants in wastewater,[6-9] river water[10-12] and, more rarely, drinking water[13-16] has been reported by many studies. Forecasts suggest that both the demographic development and the climate change will lead to higher concentrations of micropollutants in aquatic systems.[17,18]

Different water treatment processes, e.g. based on oxidation with ozone and / or adsorption onto activated carbon have been developed to improve the removal of micropollutants.[19-23] In drinking water treatment, the primary objective of such processes is the disinfection. However, during the last decades, the occurrence and the fate of organic micropollutants during these treatment processes gained increasing importance. The Landeswasserversorgung (state water supply) uses river water which is directly abstracted from the River Danube for the production of drinking water. This multi-step process is exemplary illustrated in Figure 1.
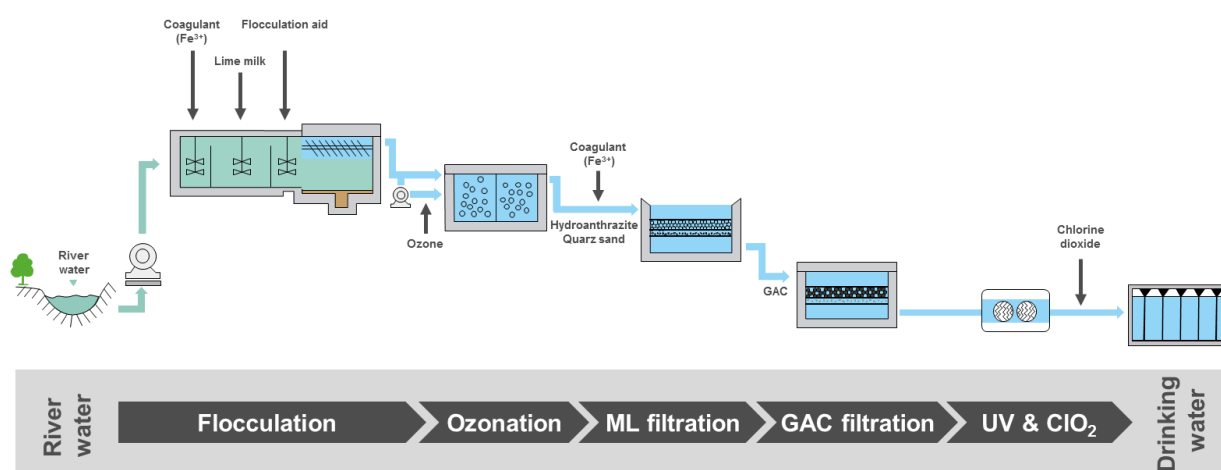

Figure 1. Processes for drinking water treatment from river water (Waterworks Langenau); ML: multi-layer, GAC: granulated activated carbon

The supply of drinking water is subjected to the strict provisions of the German drinking water ordinance (TrinkwV 2001) and the Council Directive 98/83/EC which define limit values for selected compounds. Considering the entire universe of chemicals (e.g. > 100 Mio CAS Registry Numbers[24]), however, only a very small fraction is currently regulated by the water ordinance. From the water supplier perspective, comprehensive monitoring strategies are required to deal with the variety of substances. Strong variations in quality of the raw waters require reliable treatment processes to ensure good drinking water quality. Furthermore, the behavior and the fate of micropollutants during these treatment processes (e.g. ozonation, activated carbon filtration) are of great interest. Analytical tools are required to assess the treatment efficiency for each single treatment step of the process chain.

For the monitoring of polar organic micropollutants, liquid chromatography coupled to electrospray ionization triple quadrupole tandem mass spectrometry (LC-ESI-MS/MS) is the method of choice in environmental trace analysis.[25-28] The high selectivity achieved by selected reaction monitoring in combination with good sensitivity enables the analysis at trace level. In most instances, pre-concentration techniques (e.g. solid phase extraction) are no longer necessary to achieve prescribed limits of quantification.[29-32] These analyses, however, are restricted to known target compounds and require reference standards during the method set-up. Due to the coupling with chromatographic techniques, the number of acquired transitions per run is limited to, realistically, less than 150.[33,34] With regard to the monitoring of water resources, this represents a serious restriction as only a very limited number of known and available compounds can be monitored. Such approaches usually fail to monitor unknown or unexpected components.[35] For the assessment of water treatment processes, the elimination rates of present target compounds are normally quantified.[21,22] By such approaches, however, the complete evaluation of the treatment process is based on a small number of components.[36,37] Assuming many transformation products to be unknown yet, this group is particularly underrepresented by common analytical techniques.[38-40] These restrictions expose the weaknesses of targeted approaches and emphasize the demand for new strategies.

Recent developments in high-resolution mass spectrometry coupled with liquid chromatography (LC-HRMS) have initiated new possibilities for the analysis of micropollutants without having any a priori information available.[41-43] Modern HRMS instruments provide accurate mass data while combining sufficient selectivity and sensitivity for the determination of trace substances in complex environmental matrices.[44-46] LC-HRMS has emerged as a powerful analytical tool as it enables the detection of thousands of compounds within a single run and does not require a reference standard during method set-up. The information of all detectable compounds, including unknown or unexpected substances, is available from the LC-HRMS data and can be used for a more comprehensive monitoring. Comparing the changes in the signal pattern of all detectable compounds in the influent and effluent provides a more holistic picture of the treatment process and thus allows a more reliable assessment of its treatment efficiency.[47-52]

While the analytical technique LC-HRMS itself was dynamically improved over the last years, the software algorithms[53,54] to reliably process the wealth of acquired data remained below expectations, although there were notable improvements.[55-60] The insufficient reliability of the data, mainly caused by data processing, is the bottleneck in these untargeted approaches.

2 OBJECTIVES

From the water supplier's perspective, the comprehensive monitoring of raw and drinking water is important as micropollutants might affect the quality of these resources. Untargeted approaches also allow the monitoring of unknown or unexpected compounds and thus enable timely detection of possible risks for the water supply. Furthermore, the fate and the behavior of micropollutants during water treatment processes (e.g. ozonation, activated carbon filtration) is of great interest to assess the performance of these processes based on all detectable information. The monitoring of processes or the optimization of different operating conditions are important fields of application.

In non-target screening, several thousands of signals are normally detectable within a single sample making manual review no longer a reasonable option. Instead, sophisticated algorithms are needed to automatically process the wealth of data. It was found that the processing of LC-HRMS data is a critical step which is prone to the generation of false positive and false negative findings.[55,56]

This thesis addresses the following main objectives:

- o Illustrate the problems in peak recognition particularly for low abundant signals and develop strategies to reduce the number of false positive / negative findings
- o Develop data processing strategies particularly aimed at assessing water treatment processes using the information of all detectable compounds and estimate the reliability of the results

The work summarized in **paper 1 was aimed at showing the difficulties of false positive and false negative findings and providing general strategies to minimize these influences.**[61] Improvements caused by data processing should be underlined by comparative evaluation. The question of how many sample replicates were necessary to obtain representative results should be answered and finally, the general applicability should be evaluated critically.

The work described in the **second paper deals with the LC-HRMS data processing** which was refined with special focus on the **reliable comparison of treatment processes** based on all detectable components.[62] The consistency of the results should be underlined by developing an appropriate validation concept.

Based on the selected **case studies summarized in a book chapter,** the applicability of the developed strategies on real world applications should be demonstrated.[50] Different ways for the prioritization of generated results should be depicted and the identification of detected compounds based on novel in-silico methods should be addressed.

3 METHODS AND RESEARCH APPROACH

In the following the methods on which this thesis is based on are briefly described. More details are given in the published articles (see *Reprint of articles included in this thesis*) and the related supporting information, respectively.

Due to the large amount of inventory LC-HRMS data from the past, it was decided that changes in the existing LC-MS method are to be kept to a minimum to allow for retrospective data processing. The LC-HRMS method used for data acquisition is described in **Paper 1**. To illustrate the problems in peak recognition on the basis of a systematic evaluation, multiple technical replicates (i.e. repeated measurements of the same sample) were subjected to peak finding and peak alignment. For each detected signal, the rate of recognition was determined by counting its occurrences (i.e. how often was the feature recognized) relative to the total number of technical replicates. Finally, the mean rate of recognition ($\overline{RR}$) was calculated across all features. To underline the problematics in peak recognition, the $\overline{RR}$ was calculated from data which were processed using different intensity thresholds. In total, four different matrices spiked with varying concentrations of up to 263 target compounds were investigated. Spiked targets were treated as unknowns, extracted from the peaks table (based on mass and retention time) and evaluated separately. A key point for minimizing the false positive detections while still retaining real peaks of interest was the post-acquisition filtering of extracted ion chromatograms (EIC). After peak integration, various peak characteristics (e.g. peak width) were used to filter the peak lists for false positive detections. In addition to lower and upper cutoffs, also the fluctuation across technical replicates was considered for the design of valuable filter criteria which allow distinguishing between real peaks and noise. During data processing, five processing models, requiring different numbers of technical replicates (e.g. duplicates, triplicates), were introduced. The data sets were split into unknown signals and signals caused by the spiked target compounds. For the evaluation of the five processing models, a combinatorial approach was conducted in order to compare all possible combinations of sample replicates. The rates of recognition were also calculated when applying the combinatorial approach for each individual model. To show the improvements resulted from data processing, the mean improvement factor ($\overline{I}$) was defined which compares the feature occurrences before and after data processing. The $\overline{RR}$ and $\overline{I}$ are meaningful characteristics which allow assessing the strengths / weaknesses of the processing models.

As proof of concept, four different matrices (ultrapure water, groundwater, river water and pretreated (secondary) wastewater) were spiked with 130 isotope-labeled standards at three concentration levels (25, 100 and 500 ng L^{-1}) with the objective to show that real peaks were barely or not removed by the developed filters. For an adequate assessment, each standard was manually reviewed before the untargeted approaches were applied and therefore allowed to evaluate the performance of the peak finding algorithm as well. This step was also required to report false positives during the automated data processing. The five models were evaluated with respect to the manual results. This was

necessary since several components were not even picked during peak finding. Such problems, however, were not caused by the EIC filtering.

The **second paper** addresses the refinement of the entire EIC filtering approach. The peak integration was performed using two fundamentally different integration algorithms. Also relationships within the data obtained by one algorithm (e.g. difference between centroid and apex retention time) as well as between the data obtained by the two algorithms (e.g. relative changes in peak heights) were considered. In total, 23 filters were applied for the EIC filtering. For parametrization of these filters, a LC-HRMS training data set was used. The manual review of all recognized signals revealed more than 1 300 true peaks. The parameters for the filters were derived from the distributions of the respective peak characteristics. Quantiles from these distributions were taken to set the minimum and maximum values (lower and upper cutoffs) for the filters. The (relative) standard deviation, (R)SD, across triplicates was taken as fluctuation criterion. The (R)SD distribution of the respective peak characteristic of all true peaks was considered to define the maximum (R)SD values (upper cutoff).

For the evaluation of the EIC filtering, recall and precision were determined for four different test data sets.[63] To accomplish this, the test data sets were manually reviewed and signals were assigned to the groups "peaks", "no peaks" or "exclude". The latter group was chosen for signals which could not unambiguously be assigned to "peaks" or "no peaks" (true / false decision is required to determine recall and precision). In addition, more than 400 target compounds were regarded throughout the individual stages of data processing.

For the assessment of water treatment processes, the changes of all detectable compounds were regarded based on the detected signal heights in the influent and effluent sample. A fold change (fc) based signal classification of all features was introduced. The pairwise sample comparison of the influent and effluent of a treatment process is fraught with pitfalls such as the threshold problematic. To bypass this, a recursive approach using the so called *Summation* algorithm was introduced. The *Summation* algorithm forces EIC integration irrespective of absence / presence of a chromatographic peak. This approach was conducted if either the influent or the effluent sample has passed all filter settings. In the final workflow, the data from the recursive signal integration are automatically generated for all features and can thus easily be retrieved without the need of reprocessing the entire data set.

To estimate the reliability of the LC-HRMS method in combination with the data processing workflow, technical replicates of influent and effluent samples of an ozonation process were compared. A "componentization" step was conducted to group isotopologues, adducts and dimers. In a combinatorial approach, all possible combinations for the pairwise comparison of sample triplicates were conducted and the final results were inspected regarding their consistency. This allows answering the question whether the same results would be obtained if the experiment was repeated. Two different aspects were considered during data evaluation:

(i) the repeatability of technical replicates within one sample, meaning influent and effluent were composed of the same sample but different replicates (e.g. Influent replicate #1 #2 #3 vs. Influent replicate #4 #5 #6). Considering this purely theoretical process, each signal should be classified to *consistency* (assuming the technical replicates to be identical). Thus, these considerations represent plausibility checks for the processing of the non-target data.

(ii) the repeatability of technical replicates between two samples, meaning influent and effluent were composed of different samples (i.e. Influent replicate #1 #2 #3 vs. Effluent replicate #1 #2 #3). Changes caused by the treatment processes should be reflected by the features assigned to different categories (e.g. *elimination*).

In addition to the mere numbers assigned to the respective categories, each feature was regarded individually, in order to preclude that the total numbers of assigned features were repeatable while the individual compositions were not. Isotope-labeled standards, well distributed across the considered retention time range, were co-injected to ensure consistent sensitivity throughout the batch and to account for matrix effects.

The real world applications discussed in the **book chapter** describe different methods for the prioritization of non-target features. In case study 1, the temporal prioritization was exemplified by a spill detection in time series data from the monitoring of river water. The search for the spill in the time series was accomplished by ranking detected features in a way that highest intensities and lowest frequencies result in high scores. Using the information of accurate mass, isotope pattern and MS² information, a possible elemental composition was generated. To link the formula to possible structures, online databases (e.g. drugbank[64], chemspider) were used. Finally, MS² information[65,66] was available for the suggested compound which matched the acquired DDA MS² spectrum leading to a level 2a identification.[67] For final confirmation, the reference standard was purchased which allowed a level 1 identification and retrospective semi-quantitative estimation of the concentrations. Multivariate statistics were used to search for other features that follow a similar time profile. A combination of principal component analysis and discriminant analysis (PCA-DA) was successfully implemented to find similarities.

In case study 2, non-target screening was applied as a forensic tool to link a groundwater contamination to a possible polluter. For prioritization of the generated data, logical connections between different samples were applied. For features which followed the stated rules, the DDA MS² spectra were exported and uploaded to the platform FOR-IDENT[68] which represents a compilation of water relevant chemicals. All structures which matched the accurate mass of the detected features (+/-10 ppm) were linked to the respective MS² spectra. The software package Metfrag[69] was applied for in-silico fragmentation of the possible structures and for the comparison to the acquired spectra. Further efforts in terms of identification were conducted for candidates with high Metfrag scores.

The third case study deals with the assessment of the fourth treatment step of a wastewater treatment plant (WWTP) using activated carbon filtration (granular activated carbon, GAC). To assess whether

or to what extent an improvement in the treatment efficiency occurred, the same influent was treated by both the conventional wastewater treatment and by the treatment comprising the additional activated carbon filter. To illustrate the differences between the two treatment branches, features were assigned to different categories (e.g. *elimination*, *formation*) and compared between the two treatment options. For features which were detected in both the influent and effluent, the changes in signal heights were considered.

4 RESULTS AND DISCUSSION

In the following, the most important results are summarized and discussed. In **paper 1**, the problematic of false positive and negative findings during untargeted peak finding were discussed.[61] Furthermore, the strengths and weaknesses of different processing strategies were assessed by comparing the repeatability across all combinations of technical replicates.

Regarding the spiked target compounds (between 100 and 1000 ng L^{-1}), the vast majority of the standards were successfully recognized. Moreover, $\overline{RR}$s of more than 95% were achieved suggesting that - in general and particularly for peaks with sufficient peak height - the peak finding algorithm works consistently. The target data obtained from the different processing models show that both the number of recognized standards and their $\overline{RR}$s were hardly affected by the applied filters. Thus, the false negative rate is not significantly increased by the various filtering approaches.

The problems in peak recognition, particularly for low abundant signals, were clearly evident if considering the $\overline{RR}$s calculated for unknown features. Contrary to target compounds, the $\overline{RR}$s resulted in substantially worse values (45% - 64%) which clearly point out the need for action. After data processing, the reported feature numbers were strongly reduced (by between 33% - 72%) while the $\overline{RR}$s were increased with respect to the unprocessed data. This suggests that many false positive candidates were removed by the filtering. The gradual increase in $\overline{RR}$ became lower the more replicates were used. An increase in $\overline{RR}$s can either be achieved by removing false positives or by correcting partially false negative findings. The latter was observed in many instances and quantified by the mean improvement factor $\overline{I}$. The $\overline{I}$s calculated for the different processing models (except for model a) clearly show the improvement with respect to the unprocessed data. It was not expected that partially false negative findings occurred to this extent. The use of replicates allows:

(i) regarding the fluctuation of the peak characteristics across the replicates as a valuable filter criterion to remove false positive findings

(ii) correcting partially false negative findings since a recognition rate of (number of replicates)$^{-1}$ is sufficient to retain such signals during data processing

These benefits, however, must be balanced against the decrease in sample throughput.

As proof of concept, four different sample matrices were spiked with 130 isotope-labeled standards (IS) at three different concentration levels. The manual peak inspection in the positive ionization mode (ESI+) revealed that, on average, 88%, 95% and 99% of all IS (123) were successfully detected at 25 ng L^{-1}, 100 ng L^{-1} and 500 ng L^{-1}, respectively. In the negative ionization mode (ESI-) considerably lower mean values of 38%, 69% and 92% of all IS (56) were detected at the same concentrations. The remaining standards were either below the limit of detection or concealed by strong matrix effects (signal suppression or interferences).[70] As has been expected, the number of recognized standards decreased with lower concentrations and higher matrix complexity. The number of IS recognized by the non-target peak finding was always smaller than the numbers manually verified. The differences

were low at the highest concentrations whereas stronger deviations were observed for the lower concentrations. False positive findings became apparent at lowest concentrations in the pretreated wastewater. The numbers of retained IS obtained by the different processing models were similar to the ones recognized by non-target peak finding. When referring the recovery rates in ESI+ to the maximum achievable number of standards (model (x) - manual), 94%, 97% and 99% were recognized at 25 ng L^{-1}, 100 ng L^{-1} and 500 ng L^{-1}, respectively. In ESI-, values of 90%, 92% and 94% were reached for the respective concentration levels. These findings show that real peaks of interest were barely or not removed by the applied filters and furthermore indicate that the peak finding algorithm is the limiting step in the processing workflow.

In ESI+, the 100 ng L^{-1} health orientation value[71,72] for drinking water can be reached for most components. More than 90% of the IS were successfully detected in ultrapure water, groundwater and river water whereas more than 80% were found in the pretreated wastewater. The negative ionization mode was generally less sensitive leading to worse recovery rates. At 100 ng L^{-1} about half of the IS were found in all matrices except for the wastewater where about one third of the IS could be detected. This indicates that in ESI- several micropollutants cannot be detected at or below the health orientation value.

In summary, the findings show that both the screening method and the data processing strategy are, albeit with some deficiencies in the negative ionization mode, generally applicable at trace level. Care must be given to complex matrices such as wastewater where insufficient selectivity and signal suppression may result in difficulties.

In the **second paper**, the LC-HRMS data processing was refined with special focus on the reliable sample comparison for the assessment of water treatment processes.[62] After parametrization of the derived filters using a training data set, the workflow was evaluated for its recall and precision.[63] The good performance of the EIC filtering is emphasized by values for precision and recall above 98% and 96%, respectively. The low numbers of false negatives were predominantly caused by features showing insufficient chromatographic separation. The false positive rate was always below 5%. In the next step, a multicomponent standard, comprising 411 target compounds at a concentration of 100 ng L^{-1}, was analyzed and processed with the EIC filtering approach. Out of 411 targets, 408 could successfully be recognized by the non-target peak finding. The three targets missed showed insufficient separation with other isobaric compounds from the mix. As the differences between the EIC apexes was smaller than the retention time tolerance used during peak alignment, only the more abundant species was reported. These problem candidates emphasize the importance of adequate chromatographic separation which is difficult to optimize in non-target screening.[47] Concerning the EIC filtering workflow, 406 out of 408 possible compounds were retained. These findings suggest that the possibility of losing real peaks of interest is deemed to be low.

For the assessment of treatment processes, the influent and effluent samples of a process were compared. In this comparative sample evaluation, also cases where one signal (either influent or effluent) falls below the intensity threshold need to be processed to determine the fold changes (fc) for peak categorization. In general two different viewpoints exist for dealing with signals below the threshold:

(i) values below the threshold are assumed to be zero (optimistic view)

(ii) values below the threshold are assumed to be the threshold (pessimistic view)

Irrespective of which view is considered to be the better option, inaccurate estimates may occur. The optimistic view tends to result in an overestimation of the groups *elimination* and *formation*, while the pessimistic view likely leads to an underestimation of these groups. Based on the applied intensity threshold and the fc intervals used for categorization, the "no influence threshold" was determined. After exceeding this threshold, the categorization is no longer affected by the choice of the viewpoint. Ideally, only features where the influent or effluent peak height exceeds this "no influence threshold" should be classified.

The distribution of the peak heights detected in training and test data revealed that more than 75% of all signals were below the "no influence threshold". Instead of discarding large parts of the data, a recursive approach was developed which allowed a more reliable assignment of signals below the intensity threshold. Here, peak heights below the intensity threshold were replaced by the median value across the triplicate peak heights determined by the *Summation* algorithm. This algorithm forces EIC integration irrespective of absence / presence of a chromatographic peak. The reported height is the difference between the data points of highest and lowest intensity within the considered summation window. In cases where no peak is present, the calculated fold change is to be seen as a signal-to-noise ratio. This strategy allows a more reliable classification of detected peaks without skewing the data by arbitrary assumptions concerning signals below the intensity threshold. The *Summation* algorithm was implemented for two different reasons:

(i) using a second algorithm for post-acquisition EIC filtering provides more valuable filtering criteria

(ii) the recursive approach is a main advantage in comparative data processing which allows a more reliable signal classification based on fold changes

The same recursive approach was applied for a less error prone blank correction. The blanks were not subjected to the EIC filtering but rather compared to signals which met the derived filtering criteria. Features were rejected if neither the influent nor the effluent satisfied the fc criteria relative to the blank.

For the evaluation of the repeatability, the stability of co-injected isotope-labeled standards (IS) was reviewed across the measured batch. The relative standard deviations calculated for the peak heights were always below 10% indicating sufficient system stability. Furthermore, the pairwise comparison of triplicates from the influent and effluent sample resulted in IS fold changes ranging between 0.88 and 1.17 which clearly fell within the interval of the category *consistency* ($0.5 \leq fc \leq 2.0$). Thus, all IS

would always be assigned to *C* (without blank correction). It was therefore assumed that sensitivity changes or matrix effects - in this case - had negligible influence on the feature classification.

The combinatorial approach was conducted for the plausibility checks (within samples comparisons) as well as for the real treatment process (between sample comparisons). For each single comparison (e.g. replicate #1 #2 #3 vs. #4 #5 #6), the number of features assigned to the five groups (E, D, C, I and F) were reported. This combinatorial approach and the obtained results are schematically illustrated in Figure 2.
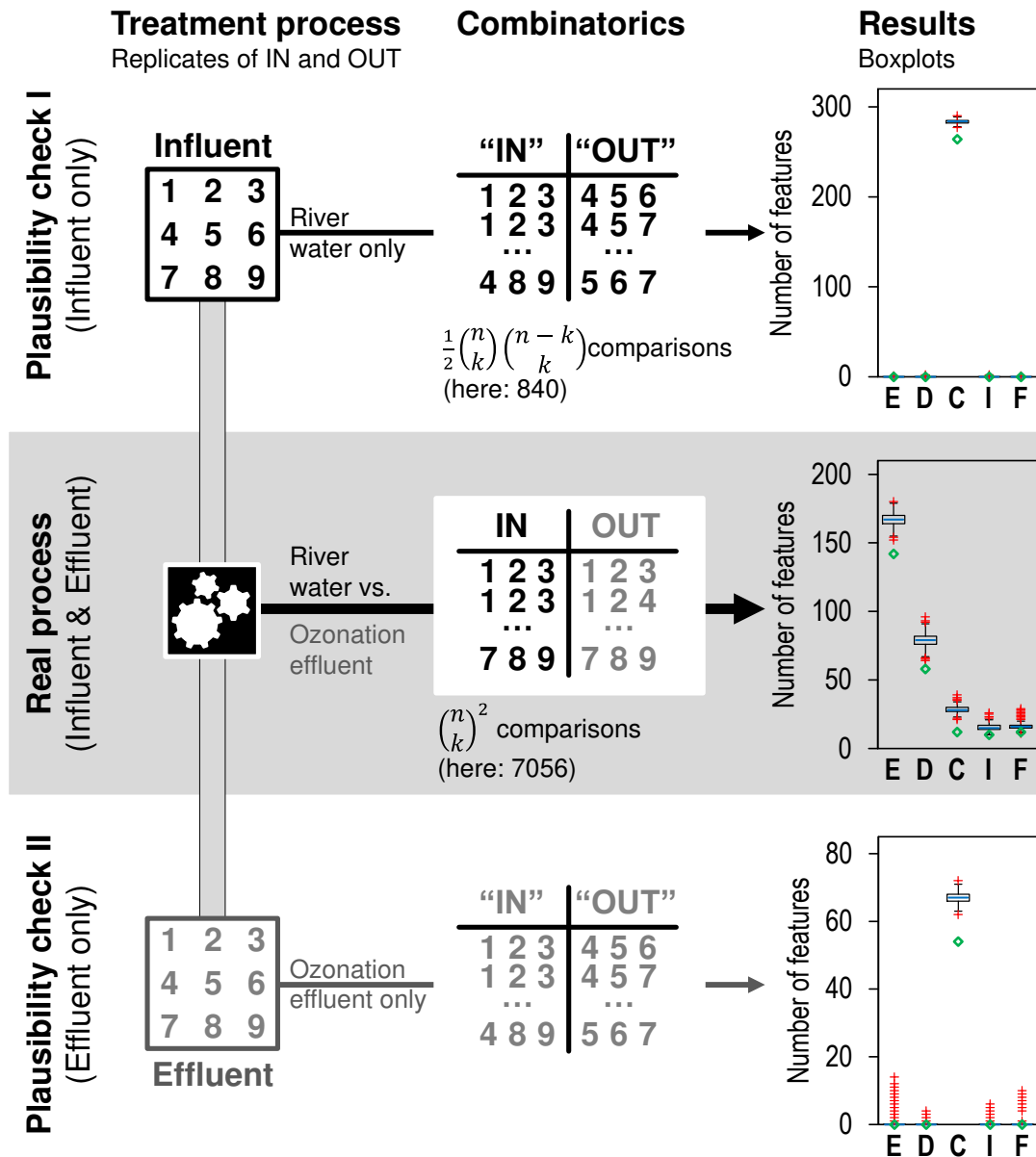
Figure 2. Combinatorial approach and results of the repeatability experiments, IN: influent sample of treatment process, OUT: effluent sample of treatment process, n: number of analyzed replicates, k: number of replicates used for data processing (here: n = 9, k = 3), E: elimination, D: decrease, C: consistency, I: increase, F: formation, features which were unambiguously (i.e. in all possible comparisons) assigned to one category are indicated by the green diamonds

However, in addition to the mere numbers, each individual feature was considered separately as the sums of all features assigned to a certain category could be repeatable while the individual compositions of the sums were not. Thus, it was also reported how often (i.e. in how many of the comparisons) a particular feature was assigned to one group. After applying the combinatorial approach, features which were ambiguously assigned to multiple categories were reviewed. The manual inspection of these problem candidates revealed three main reasons:

(1) signals were in vicinity of the intensity threshold or the blank threshold. With some signals slightly above / below these thresholds, the features were sometimes rejected during EIC filtering while other combinations of replicates passed the filters.

To overcome these problems, the data were reprocessed using less stringent criteria for the minimum value of peak area and height as well as for the blank correction which were applied to all signals detected during the first processing.

(2) the calculated fold change was close to the edge of the interval used for categorization. Across all combinations, some were slightly above / below this limit resulting in ambiguous assignment into two adjacent categories (e.g. *elimination* / *decrease*).

To bypass this problem, the algorithm was adapted to handle such cases. If the number of all possible comparisons was reached by counting two adjacent categories, the feature was unambiguously assigned to the category comprising the larger number of comparisons.

(3) real peaks showed implausible signal courses across the batch (strong fluctuations, trends). These features reflect real differences within the replicates of the same sample (unexplained circumstances) and are not subjected to data processing. A reasonable correction of this group is not possible.

The reprocessing of the data with the conducted changes mentioned in (1) and (2) resulted in a significant decrease in the number of features that were ambiguously assigned to multiple categories. These findings show that many problem candidates were merely caused by the rigid thresholds. The problem cases described in (3), however, are not correctable since these signals expose real differences between technical replicates.

As has been expected, the conducted plausibility checks indicate that differences between the technical replicates hardly exist. Almost all features were assigned to the category *consistency* while just few outliers were assigned to the remaining groups. The good repeatability of the sums of features assigned to the respective classes is indicated by the small variations of the assigned feature numbers. However, a certain offset between the number of features always assigned to *consistency* and the distributions of the sums of features assigned to this category exist. The majority of these differences were caused by implausible signal courses - see (3) - or interfering peaks.

Considering the real process (ozonation of pretreated river water), strong changes between the influent and effluent samples became apparent. Contrary to the plausibility checks, most features were assigned to the group *elimination* or *decrease*. The variation of the numbers of features assigned to the five categories throughout the different comparisons revealed low fluctuation and thus points to

sufficient repeatability. Again, an offset between the number of features always assigned to a certain category and the actual distributions was recognized for the same reasons as stated above. The treatment efficiency of the ozonation process is indicated by the large proportion of features assigned to the categories *elimination* and *decrease*. Considering the median values of the respective distributions across all conducted comparisons, 55% of all signals were assigned to *elimination*, 26% to *decrease*, 9% to *consistency*, 5% to *increase* and 5% to *formation*. More features were expected to be assigned to *formation* as the ozonation process is known for the generation of transformation products.[73-76] Even though the entire window of detectable substances is strongly enhanced by non-target screening, there are still components which cannot be captured by this method.[77] Each single step, i.e. separation, ionization, detection, as well as data processing, requires certain criteria which narrow down the diversity of substances being considered.

The application of the developed strategies to real world applications is summarized in a **book chapter**.[50] Different strategies for the prioritization (i.e. temporal, spatial and process related ones) were presented. Case study 1 exemplifies the spill detection in river water using non-target screening. The beta blocker Acebutolol was successfully identified and quantified (semi-quantitatively) in the river water. In the next step, multivariate statistics were used to search for similar time profiles which may be in conjunction with the detected spill event. After applying PCA-DA, the time profile of feature 557.255/6.6 (accurate mass / retention time) was found to have high similarities with Acebutolol. Despite further efforts in terms of identification, the feature remained a level 5 candidate.[67] This example shows the difficulties in identification if no information is available in public databases. In many cases, further analytical techniques (e.g. nuclear magnetic resonance) are required to fully elucidate the chemical structure.[78,79]

The use of non-target screening as a forensic tool in water analysis is shown in case study 2. The anthropogenic contamination of a groundwater well was recognized by target analysis (data not shown). Besides two municipal WWTPs, also an industrial WWTP is located near the groundwater well. It has been suggested that the industrial WWTP might (also) be responsible for the contamination, however, no source specific contaminations were found so far. The aim of this study was to find groundwater contaminants which can exclusively be attributed to the industrial WWTP. After data processing, including all filtering steps, between 1 400 and 2 400 features remained in the different peak tables. Based on the analytical request, logical connections were applied between the five different samples. The complete list of thousands of features could be reduced to 54 candidates that might help answering the initial question. Further prioritization was accomplished by in-silico fragmentation of possible structures linked to the accurate feature masses. Seven candidates were ranked with high scores and therefore considered in more details. The component Triphenylphosphine oxide (TPPO)[80] was successfully identified as a level 1 candidate. Further compounds related to TPPO were identified as level 2a and 2b candidates.[67] The contamination with TPPO is much likely related to

the Wittig reaction where it is formed as a byproduct.[11,81] The determined TPPO concentrations in the groundwater well were higher than 5 µg L^{-1} and thus clearly above the health orientation values of 0.1 µg L^{-1}.[71,72] These findings clearly indicate the industrial influence on the regarded groundwater well and thus answer the initial question. Further measures were initiated to reduce the TPPO discharge from the industrial WWTP.

Case study 3 was aimed at assessing the purification efficiency of the fourth treatment step of a WWTP in comparison to the conventional wastewater treatment. Based on the changes throughout the treatment, the signals were classified to different classes which allowed a comparison of the two treatment options. The higher purification performance using the GAC filtration was already evident if comparing the effluents of both treatment options. The fourth treatment step showed a reduction of more than 14% in terms of the feature numbers. The number of features assigned to the category elimination was 42% higher for the GAC filtration. Considering the group *formation*, 15% more features were formed in the conventional treatment. Features detected in both the influent and effluent sample were classified based on the relative changes in peak height. The proportion of features eliminated by at least 60% was three times higher for the GAC treatment. In the conventional treatment, 75% more features were subjected to increasing intensities.

This simple comparison of the influent and effluent samples clearly indicated a lower rate of newly formed features and at the same time a higher elimination for features undergoing the additional GAC filtering step. A closer look at the features occurring in both the influent and effluent samples revealed higher numbers of features with decreasing signal intensities in the GAC filter compared to the conventional treatment. As was to be expected, the benefit of this additional adsorption step (GAC filter) is clearly evident. With regard to organic trace substances, the implementation of such an additional filtering step in wastewater treatment is to be recommended.

5 CONCLUSION AND FUTURE NEEDS

In this thesis, LC-HRMS data processing strategies for the assessment of water treatment processes were developed. Improvement of the data quality, i.e. minimizing of false positive and false negative findings, was a key point. On this basis, sophisticated workflows for the reliable sample comparison were developed and validated using novel concepts.

It was found that triplicate analysis and the post-acquisition filtering of extracted ion chromatograms significantly improved both the repeatability and the data quality. These findings suggest that problems were primarily caused by insufficient processing rather than by measurement problems. It was further found that the peak finding is the limiting step of the entire workflow. The following improvements are therefore suggested: (i) combining different peak finding algorithms to pool their individual strengths and thus increase the feature coverage (ii) replace the fixed intensity threshold by a signal-to-noise based peak recognition.

For the assessment of water treatment processes, the reliability of the data needs to be high to avoid skewing of results. The recursive peak integration was found to be a major improvement for the reliable fold change categorization of the signals. The use of internal standards is strongly recommended to estimate the influence of matrix effects, which have to be seen with particular caution. The strength of the entire LC-HRMS processing strategy is indicated by the good repeatability of the signal classification. The plausibility checks considered during data processing fulfilled the expectations and emphasized the reliability of the results. Further efforts need to be conducted to reduce implausible signal courses and redundancies in data interpretation. In the presented workflow(s), the componentization is conducted after peak alignment and hence requires much higher retention time and mass tolerances. It is therefore recommended to perform this grouping step before peak alignment as smaller tolerances would decrease the number of false positive linkages. The assignment of in-source fragmentation is another point which requires further research needs. Considering the MS² information (DDA or DIA, e.g. Swath^{TM}), potential in-source fragments could be linked to parent compounds. Despite these restrictions, non-target screening using LC-HRMS is a very powerful tool for the assessment of water treatment processes. The much broader view will help to reduce existing knowledge gaps and to identify new components of interest.

The real world applications underline the general applicability of non-target screening in combination with sophisticated workflows and prioritization concepts. The identification of unknown or unexpected chemicals represents the bottleneck of non-target screening workflows and emphasizes further research needs.

The LC-HRMS opens up a new dimension for monitoring the fate of organic contaminants in the aquatic environment and also during treatment processes. The application of the presented concepts is not limited to treatment processes. The processing strategies and the combinatorial validation concept may be transferred to other research fields where sample comparisons are conducted.

6 ACKNOWLEDGEMENT

There are numerous people who supported me and my work in the last years. I would like to express my sincere gratitude to the following persons:

o Klaus Kümmerer for the support of my thesis from the university side, the helpful discussions of results and valuable feedback concerning the improvement of research articles.

o Wolfgang Schulz for the steady support and guidance during this work. He was always willing to share his expertise and never reluctant to help. Some discussions intended to last few minutes turned into debates lasting for hours (even after 6 pm). Never forcing, always suggesting - he allowed this dissertation to be my own work and steered me in the right direction when necessary. His advice and encouragement have been immensely valuable and his network within the water community greatly helped to exchange experiences with other specialists.

o Christian Zwiener and Wolf von Tümpling for spending their precious time for being reviewers of this thesis.

o Rudi Winzenbacher for providing the opportunity and environment for this dissertation which, surely, are above standard. The practice oriented research on such interesting questions with importance to the water supply was very motivating.

o Thomas Lucke for many fruitful discussions concerning the development of processing strategies. His feedback from the practical application of the strategies was extremely helpful for improving the algorithms and workflows.

o Wolfram Seitz for his quick response and edits to anything I sent his way, his help with organizational matters and discussions on various contributions.

o Marina Bischoff for her assistance in the laboratory and operation of the LC-HRMS system. Her precise and systematic way of working was a great help during this work.

o Lena Stütz and Regine Fischeder for proofreading of the articles and providing helpful suggestions for improving the content.

o Personally, I would like to thank my family for their support throughout the years of the studies and this dissertation: financially, practically and with moral support. A special acknowledgement goes to my girlfriend Katrin, who encouraged and supported me during the studies and this dissertation.

7 REFERENCES

(1) Schwarzenbach, R. P.; Escher, B. I.; Fenner, K.; Hofstetter, T. B.; Johnson, C. A.; von Gunten, U.; Wehrli, B. *Science* **2006**, *313*, 1072-1077.

(2) Richardson, S. D.; Ternes, T. A. *Anal. Chem.* **2014**, *86*, 2813-2848.

(3) Howard, P. H.; Muir, D. C. G. *Environ. Sci. Technol.* **2010**, *44*, 2277-2285.

(4) Reemtsma, T.; Weiss, S.; Mueller, J.; Petrovic, M.; Gonzalez, S.; Barcelo, D.; Ventura, F.; Knepper, T. P. *Environ. Sci. Technol.* **2006**, *40*, 5451-5458.

(5) Kümmerer, K. *J. Environ. Econ. Manag.* **2009**, *90*, 2354-2366.

(6) Seitz, W.; Winzenbacher, R. *Environ Monit Assess* **2017**, *189*, 244.

(7) Loos, R.; Carvalho, R.; António, D. C.; Comero, S.; Locoro, G.; Tavazzi, S.; Paracchini, B.; Ghiani, M.; Lettieri, T.; Blaha, L.; Jarosova, B.; Voorspoels, S.; Servaes, K.; Haglund, P.; Fick, J.; Lindberg, R. H.; Schwesig, D.; Gawlik, B. M. *Water Res.* **2013**, *47*, 6475-6487.

(8) Castronovo, S.; Wick, A.; Scheurer, M.; Nödler, K.; Schulz, M.; Ternes, T. A. *Water Res.* **2017**, *110*, 342-353.

(9) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. *Environ. Pollut.* **2014**, *184*, 25-32.

(10) Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P. *Water Res.* **2015**, *87*, 145-154.

(11) Schlüsener, M. P.; Kunkel, U.; Ternes, T. A. *Environ. Sci. Technol.* **2015**, *49*, 14282-14291.

(12) Loos, R.; Gawlik, B. M.; Locoro, G.; Rimaviciute, E.; Contini, S.; Bidoglio, G. *Environ. Pollut.* **2009**, *157*, 561-568.

(13) Fleischer, S.; Weiss, S. C.; Lucke, T.; Seitz, W.; Schulz, W.; Weber, W. H. *Ozone: Sci. Eng.* **2015**, *37*, 441-449.

(14) Zahn, D.; Frömel, T.; Knepper, T. P. *Water Res.* **2016**, *101*, 292-299.

(15) Herrmann, M.; Menz, J.; Olsson, O.; Kümmerer, K. *Water Res.* **2015**, *85*, 11-21.

(16) Segura, P. A.; MacLeod, S. L.; Lemoine, P.; Sauvé, S.; Gagnon, C. *Chemosphere* **2011**, *84*, 1085-1094.

(17) Delpla, I.; Jung, A. V.; Baures, E.; Clement, M.; Thomas, O. *Environ. Int.* **2009**, *35*, 1225-1233.

(18) Vatovec, C.; Phillips, P.; Van Wagoner, E.; Scott, T.-M.; Furlong, E. *Sci. Total Environ.* **2016**, *572*, 906-914.

(19) Yang, Y.; Ok, Y. S.; Kim, K.-H.; Kwon, E. E.; Tsang, Y. F. *Sci. Total Environ.* **2017**, *596*, 303-320.

(20) Prasse, C.; Stalter, D.; Schulte-Oehlmann, U.; Oehlmann, J.; Ternes, T. A. *Water Res.* **2015**, *87*, 237-270.

(21) Knopp, G.; Prasse, C.; Ternes, T. A.; Cornel, P. *Water Res.* **2016**, *100*, 580-592.

(22) Margot, J.; Kienle, C.; Magnet, A.; Weil, M.; Rossi, L.; de Alencastro, L. F.; Abegglen, C.; Thonney, D.; Chèvre, N.; Schärer, M.; Barry, D. A. *Sci. Total Environ.* **2013**, *461-462*, 480-498.

(23) Luo, Y.; Guo, W.; Ngo, H. H.; Nghiem, L. D.; Hai, F. I.; Zhang, J.; Liang, S.; Wang, X. C. *Sci. Total Environ.* **2014**, *473*, 619-641.

(24) Media Releases **2015**. https://www.cas.org/news/media-releases/100-millionth-substance (accessed Sep 29, 2017)

(25) Villagrasa, M.; López de Alda, M.; Barceló, D. *Anal. Bioanal. Chem.* **2006**, *386*, 953-972.

(26) Pérez-Fernández, V.; Mainero Rocca, L.; Tomai, P.; Fanali, S.; Gentili, A. *Anal. Chim. Acta* **2017**, *983*, 9-41.

(27) Sancho, J. V.; Pozo, O. J.; Hernández, F. *Analyst* **2004**, *129*, 38-44.

(28) Pozo, Ó. J.; Sancho, J. V.; Ibáñez, M.; Hernández, F.; Niessen, W. M. A. *TrAC-Trend. Anal. Chem.* **2006**, *25*, 1030-1042.

(29) Berset, J.-D.; Brenneisen, R.; Mathieu, C. *Chemosphere* **2010**, *81*, 859-866.

(30) Reemtsma, T.; Alder, L.; Banasiak, U. *J. Chromatogr. A* **2013**, *1271*, 95-104.

(31) Seitz, W.; Schulz, W.; Weber, W. H. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 2281-2285.

(32) Barceló, D.; Petrovic, M. *TrAC-Trend. Anal. Chem.* **2007**, *26*, 2-11.

(33) Ferrer, I.; Thurman, E. M.; Zweigenbaum, J. A. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 3869-3882.

(34) Huntscha, S.; Singer, H. P.; McArdell, C. S.; Frank, C. E.; Hollender, J. *J. Chromatogr. A* **2012**, *1268*, 74-83.

(35) Tengstrand, E.; Rosén, J.; Hellenäs, K.-E.; Åberg, K. M. *Anal. Bioanal. Chem.* **2013**, *405*, 1237-1243.

(36) Jekel, M.; Dott, W.; Bergmann, A.; Dünnbier, U.; Gnirss, R.; Haist-Gulde, B.; Hamscher, G.; Letzel, M.; Licha, T.; Lyko, S.; Miehe, U.; Sacher, F.; Scheurer, M.; Schmidt, C. K.; Reemtsma, T.; Ruhl, A. S. *Chemosphere* **2015**, *125*, 155-167.

(37) Moschet, C.; Wittmer, I.; Simovic, J.; Junghans, M.; Piazzoli, A.; Singer, H.; Stamm, C.; Leu, C.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 5423-5432.

(38) Petrie, B.; Barden, R.; Kasprzyk-Hordern, B. *Water Res.* **2015**, *72*, 3-27.

(39) Menz, J.; Toolaram, A. P.; Rastogi, T.; Leder, C.; Olsson, O.; Kümmerer, K.; Schneider, M. *Environ. Int.* **2017**, *98*, 171-180.

(40) Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. *Anal. Chem.* **2015**, *87*, 12121-12129.

(41) Moschet, C.; Piazzoli, A.; Singer, H.; Hollender, J. *Anal. Chem.* **2013**, *85*, 10312-10320.

(42) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M.; Onghena, M.; Kunkel, U.; Schulz, W., et al. *Anal. Bioanal. Chem.* **2015**, *407*, 6237-6255.

(43) Müller, A.; Schulz, W.; Ruck, W. K. L.; Weber, W. H. *Chemosphere* **2011**, *85*, 1211-1219.

(44) Krauss, M.; Singer, H.; Hollender, J. *Anal Bioanal Chem* **2010**, *397*, 943-951.

(45) Hernández, F.; Sancho, J. V.; Ibáñez, M.; Abad, E.; Portolés, T.; Mattioli, L. *Anal. Bioanal. Chem.* **2012**, *403*, 1251-1264.

(46) Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. *Environ. Sci. Technol.* **2015**, *49*, 12333-12341.

(47) Nürenberg, G.; Schulz, M.; Kunkel, U.; Ternes, T. A. *J. Chromatogr. A* **2015**, *1426*, 77-90.

(48) Merel, S.; Anumol, T.; Park, M.; Snyder, S. A. *J. Hazard. Mater.* **2015**, *282*, 75-85.

(49) Parry, E.; Young, T. M. *Water Res.* **2016**, *104*, 72-81.

(50) Bader, T.; Schulz, W.; Lucke, T.; Seitz, W.; Winzenbacher, R. In: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening − Strategies and Workflows Volume 2*; American Chemical Society, **2016**, 49-70.

(51) Heuett, N. V. Dissertation, Florida International University, **2015**, http://digitalcommons.fiu.edu /etd/2194.

(52) Toedtli, L. Master thesis, ETH Zurich and eawag, **2015**, http://library.eawag.ch/EAWAG-Publications/openaccess/Diplomarbeiten/UCHEM/diplomarbeit_lauratoedtli.pdf.

(53) Katajamaa, M.; Orešič, M. *J. Chromatogr. A* **2007**, *1158*, 318-328.

(54) Zhang, J.; Gonzalez, E.; Hestilow, T.; Haskins, W.; Huang, Y. *Curr. Genom.* **2009**, *10*, 388-401.

(55) Vergeynst, L.; Van Langenhove, H.; Demeestere, K. *Anal. Chem.* **2015**, *87*, 2170-2177.

(56) Coble, J. B.; Fraga, C. G. *J. Chromatogr. A* **2014**, *1358*, 155-164.

(57) Yu, W.; He, Z.; Liu, J.; Zhao, H. *J. Proteome Res.* **2008**, *7*, 123-129.

(58) Yu, T.; Jones, D. P. *Bioinformatics* **2014**, *30*, 2941-2948.

(59) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779-787.

(60) Brodsky, L.; Moussaieff, A.; Shahaf, N.; Aharoni, A.; Rogachev, I. *Anal. Chem.* **2010**, *82*, 9177-9187.

(61) Bader, T.; Schulz, W.; Kümmerer, K.; Winzenbacher, R. *Anal. Chim. Acta* **2016**, *935*, 173-186.

(62) Bader, T.; Schulz, W.; Kümmerer, K.; Winzenbacher, R. *Anal. Chem.* **2017**, *89*, 13219-13226.

(63) Fawcett, T. *Pattern Recognit. Lett* **2006**, *27*, 861-874.

(64) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. *Nucleic Acids Res.* **2006**, *34*, D668-672.

(65) HighChem. LLC, H., Ed.: Slovakia, Advanced Mass Spectral Database, www.mzcloud.org (accessed Apr 15, 2016).

(66) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N., et al. *J. Mass Spectrom.* **2010**, *45*, 703-714.

(67) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 2097-2098.

(68) FOR-IDENT, https://water.for-ident.org, (accessed Dec 01, 2015, login only)

(69) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. *J. Cheminf.* **2016**, *8*, 1-16.

(70) Kruve, A.; Herodes, K.; Leito, I. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1159-1168.

(71) Bergmann, A.; Weber, F.-A.; Meiners, H. G.; Müller, F. *Environ. Sci. Eur* **2014**, *26*, 1-14.

(72) Umweltbundesamt. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* **2003**, *46*, 249-251.

(73) Hübner, U.; von Gunten, U.; Jekel, M. *Water Res.* **2015**, *68*, 150-170.

(74) De Vera, G. A.; Stalter, D.; Gernjak, W.; Weinberg, H. S.; Keller, J.; Farré, M. J. *Water Res.* **2015**, *87*, 49-58.

(75) Postigo, C.; Richardson, S. D. *J. Hazard. Mater.* **2014**, *279*, 461-475.

(76) Merel, S.; Zwiener, C. In: *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening − Strategies and Workflows Volume 2*; American Chemical Society, **2016**, 3-27.

(77) Reemtsma, T.; Berger, U.; Arp, H. P. H.; Gallard, H.; Knepper, T. P.; Neumann, M.; Quintana, J. B.; Voogt, P. d. *Environ. Sci. Technol.* **2016**, *50*, 10308-10315.

(78) Wendel, F. M.; Lütke Eversloh, C.; Machek, E. J.; Duirk, S. E.; Plewa, M. J.; Richardson, S. D.; Ternes, T. A. *Environ. Sci. Technol.* **2014**, *48*, 12689-12697.

(79) Brezina, E.; Prasse, C.; Wagner, M.; Ternes, T. A. *Environ. Sci. Technol.* **2015**, *49*, 10449-10456.

(80) Knepper, T. P.; Karrenbrock, F. In: *The Rhine*, Knepper, T. P., Ed.; Springer: Berlin, D, **2006**, 235-254.

(81) Maryanoff, B. E.; Reitz, A. B. *Chem. Rev.* **1989**, *89*, 863-927.

APPENDIX

Curriculum vitae

Tobias Bader, born August 14th, 1989 in Weißenhorn, Germany

Education

Oct 2014 - present	**Ph.D. candidate (Dr. rer. nat.)**
	Institute of Sustainable and Environmental Chemistry,
	Faculty of Sustainability Sciences, Leuphana University Lüneburg, Germany
Oct 2012 - Feb 2014	**Master of Science in Analytical and Bioanalytical Chemistry**
	Aalen University of Applied Sciences, Germany
Oct 2008 - Aug 2012	**Bachelor of Science in Chemistry**
	Aalen University of Applied Sciences, Germany
Sep 2006 - Aug 2008	**Vocational diploma**
	Technical Secondary School Neu-Ulm, Germany

Professional Experience

Mar 2014 - present	**Research fellow**
	Zweckverband Landeswasserversorgung, Langenau, Germany
	Laboratory for operation control and research
	○ Dissertation "Mining of LC-HRMS data for the assessment of water treatment processes" (in cooperation with the Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg)
	○ Supervision of B.Sc. and M.Sc. students
	○ Working on research projects
Sep 2013 - Feb 2014	**Master Thesis**
	Health Canada, ON, Canada
	"Characterization of biopharmaceutical final formulations by combining separation with on-line identification by mass spectrometry: Interferon alpha-2"
Nov 2012 - Mar 2013	**Research project work**
	Aalen University of Applied Sciences, Germany
	○ "Determination of metal complexes of bleocin by CE-MS and MALDI-TOF/TOF MS"
	○ "Development of a method for various metals and boron in matrices containing hydrofluoric acid and silicon by ICP-OES"

Feb 2012 - Aug 2012	**Bachelor Thesis**
	Aalen University of Applied Sciences, Germany
	"Method development for separation of anions in an acid background-electrolyte and the application-related analysis of the glucosinolate pattern in phloem saps by capillary zone electrophoresis - mass spectrometry"
Aug 2010 - Feb 2011	**Full-time internship**
	Zweckverband Landeswasserversorgung, Langenau, Germany
	"Method development and analysis of trace substances by HPLC-MS/MS"

Publications

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2016**). General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta* 935: 173-186. DOI: 10.1016/j.aca.2016.06.030

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2017**). LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes. *Analytical Chemistry* 89: 13219-13226. DOI: 10.1021/acs.analchem.7b03037

Tobias Bader, Wolfgang Schulz, Thomas Lucke, Wolfram Seitz, Rudi Winzenbacher (**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* (Vol. 1242, pp. 49-70): American Chemical Society. DOI: 10.1021/bk-2016-1242.ch003

Tobias Bader, Wolfgang Schulz, Thomas Lucke, Rudi Winzenbacher, Klaus Kümmerer (**2018**). Weitergehende Prozessbewertung mittels Non-Target-Screening bei der Landeswasserversorgung. gwf-Wasser|Abwasser 02|2018 48-51

Conference contributions

Jahrestagung der Wasserchemischen Gesellschaft **2015**, May 11 - 13, Schwerin (DE)
"Praktische Erfahrungen bei der Roh- und Trinkwasserüberwachung mittels HPLC-HRMS im Wasserwerk Langenau" (Presentation)

Nontarget2016 Conference **2016**, May 29 - June 03, Monte Verità, Ascona (CH)
"General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry" (Presentation)

21st International Mass Spectrometry Conference **2016**, August 20 - 26, Toronto (CA)
"General strategies to increase the repeatability in environmental non-target screening by liquid chromatography-high resolution mass spectrometry" (Presentation)

Jahrestagung der Wasserchemischen Gesellschaft **2016**, May 02 - 04, Bamberg (DE)
"Zuordnung von Kontaminationsquellen mittels Non-Target-Screening" (Poster)

Jahrestagung der Wasserchemischen Gesellschaft **2017**, May 22 - 24, Donaueschingen (DE)
"Bewertung von Aufbereitungsprozessen mittels Non-Target-Screening - Validitätsprobleme durch Matrixeffekte?" (Poster and flash talk)

23rd International Symposium on Separation Sciences **2017**, September 19 - 22, Vienna (AT)
"Non-target screening for the evaluation of water treatment processes" (Poster)

Jahrestagung der Wasserchemischen Gesellschaft **2018**, May 07 - 09, Papenburg (DE)
"Bewertung von Aufbereitungsprozessen mittels Non-Target-Screening - Validitätsprüfung anhand der Langenauer Donauwasseraufbereitung" (Poster and flash talk)

Reprint of articles included in this thesis

Paper 1:

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2016**). General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta* 935: 173-186. DOI: 10.1016/j.aca.2016.06.030

Paper 2:

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher (**2017**). LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes. *Analytical Chemistry* 89: 13219-13226. DOI: 10.1021/acs.analchem.7b03037

Chapter 1:

Tobias Bader, Wolfgang Schulz, Thomas Lucke, Wolfram Seitz, Rudi Winzenbacher (**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* (Vol. 1242, pp. 49-70): American Chemical Society. DOI: 10.1021/bk-2016-1242.ch003

PAPER 1

General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher

General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry

Tobias Bader [a,b,*], Wolfgang Schulz [a], Klaus Kümmerer [b], Rudi Winzenbacher [a]

[a] Laboratory for Operation Control and Research, Zweckverband Landeswasserversorgung, Am Spitzigen Berg 1, 89129 Langenau, Germany
[b] Sustainable Chemistry and Material Resources, Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg, Scharnhorststraße 1/C13, 21335 Lüneburg, Germany

HIGHLIGHTS

- Combinatorial approach for validation of the data evaluation in non-target screening.
- Replicates decrease false negative and false positive findings.
- Signal fluctuations emerged as powerful filter criteria.
- Data processing increases repeatability.
- Screening method and data evaluation in general applicable at trace levels.

GRAPHICAL ABSTRACT
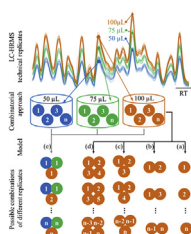
ARTICLE INFO

ABSTRACT

This article focuses on the data evaluation of non-target high-resolution LC-MS profiles of water samples. Taking into account multiple technical replicates, the difficulties in peak recognition and the related problems of false positive and false negative findings are systematically demonstrated. On the basis of a combinatorial approach, different models involving sophisticated workflows are evaluated, particularly with regard to the repeatability. In addition, the improvement resulting from data processing was systematically taken into consideration where the recovery of spiked standards emphasized that real peaks of interest were barely or not removed by the derived filter criteria. The comprehensive evaluation included different matrix types spiked with up to 263 analytical standards which were analyzed repeatedly leading to a total number of more than 250 injections that were incorporated in the assessment of different models of data processing. It was found that the analysis of multiple replicates is the key factor as, on the one hand, it provides the option of integrating valuable filters in order to minimize the false positive rate and, on the other hand, allows correcting partially false negative findings occurring during the peak recognition. The developed processing strategies including replicates clearly point to an enhanced data quality since both the repeatability as well as the peak recognition could be considerably improved. As proof of concept, four different matrix types, including a wastewater treatment plant (WWTP) effluent, were spiked with 130 isotopically labeled standards at different concentration levels. Despite the stringent filter criteria, at 100 ng L^{-1} recovery rates of up to 93% were reached

in the positive ionization mode. The proposed model, comprising three technical replicates, filters less than 5% and 2% of the standards recognized at 100 and 500 ng L^{-1}, respectively and thus indicates the general applicability of the presented strategies.

1. Introduction

Recent developments in high-resolution mass spectrometry coupled to liquid chromatography (LC-HRMS) have initiated new possibilities for the analysis of micropollutants without having any a priori information available [1]. Modern HRMS instruments provide accurate mass data while combining sufficient selectivity and sensitivity, which allows the determination of trace substances in complex environmental matrices [2–5]. LC-HRMS has emerged as a powerful tool as it enables the detection of thousands of compounds within a single run and does not require a reference standard during the method set-up [1,6] as it is usual for triple quadrupole instruments. In contrast, conventional targeted analytical methods only allow to monitor a tiny fraction of known contaminants per run and therefore miss unknowns, such as transformation products and chemicals which were initially not anticipated regardless of how high their concentrations might be [7]. While the untargeted data acquisition offers a variety of advantages, sophisticated processing strategies are needed to handle the wealth of data and to extract the significant information. In the first step of most untargeted workflows, peak finding algorithms are used to extract peaks (features) from the existing data set [8,9]. Throughout a large number of studies, it became evident that the peak finding step, independently of the software used, reveals type I (false positive) and type II (false negative) errors. The intensity threshold to differentiate noise from real peaks is one of the most important parameters in non-target screening. Setting a low intensity threshold favors type I errors whereas higher thresholds will cause real peaks of interest to be missed. Many authors found that false positives, i.e. noise or matrix background recorded as peaks, as well as false negatives, i.e. true peaks which were not recognized, represent the main challenge especially if dealing with low abundant signals [6,10,11]. Consequently, different strategies have been developed in order to distinguish real peaks from noise signals while still keeping a chosen set of known peaks, i.e. spiked standards [7,12–15]. Many approaches attempt to emphasize the temporal, spatial or process-based variances among different samples [16,17] using multivariate statistics. This, however, requires that the variance is attributed to true differences between the samples [14]. Problems in peak recognition between different samples appear largely random and therefore hamper the elucidation of discriminating features as real differences are superimposed by apparent ones created during the peak finding.

The objective of this study was to illustrate the problematics of false positive and false negative findings on the basis of a systematic evaluation of multiple technical replicates of different matrix types. The improvement in feature detection by comparative evaluation of different processing strategies involving various models is shown. The features' frequency of recognition was therefore adopted as a measure for the repeatability of the method, which was determined using a combinatorial approach that is exemplified in the supplementary video. Starting with very poor repeatability, pointing to the above mentioned problems of type I and type II errors, the rate of recognition is successfully increased by applying different models and the involved filter criteria. Moreover, the question of how many replicates are needed to

provide a representative result will be discussed.

Supplementary video related to this article can be found at http://dx.doi.org/10.1016/j.aca.2016.06.030.

2. Materials and methods

2.1. Chemicals

The standard substances combined in the stock solutions (supplemental table S1 and S2) were purchased from various suppliers (supplemental table S3) Isotopically labeled standards (supplemental table S4) were provided by the Swiss Federal Institute for Aquatic Science and Technology (Dübendorf, Switzerland) as a multi-component standard. Water (Rotisolv® Ultra LC-MS), acetonitrile (Rotisolv® ≥ 99.95%, LC-MS-Grade) and methanol (Rotisolv® ≥ 99.95%, LC-MS-Grade) were purchased from Carl Roth (Karlsruhe, Germany) while formic acid (for mass spectrometry ~ 98%) was supplied by Sigma-Aldrich (Steinheim am Albuch, Germany). For mass calibration, the APCI positive and negative calibration solutions were delivered by Sciex (Framingham, USA).

2.2. Sample preparation

Except spiking the samples with different stock solutions, in general, no prior sample preparation was performed to avoid discriminating against certain substances. In non-target analysis, we believe this to be very important even though the analytical method might suffer from the lack of pretreatment. Samples that obviously contained suspended matter were centrifuged for 5 min at 5000 rpm before the analyses.

2.2.1. Stock solutions

An exact amount of each pure substance was dissolved in methanol, stored at −18 °C and, shortly before the analyses, multi component standards were produced by diluting numerous standard solutions to the required concentration. Based on this procedure, stock solution I (supplemental table S1) covering 32 pharmaceutical drugs as well as stock solution II (supplemental table S2) comprising 263 various substances from different classes (e.g. pesticides, biocides, industrial chemicals and corrosion inhibitors) were produced. Stock solution III (supplemental table S4) containing 130 isotopically labeled substances was received fully prepared and thus only had to be diluted to the desired concentration. Note that reference substances were selected to be compatible with the method applied whereas special attention was given to adequately cover a relevant polarity range (logP ≈ −1 to 5).

2.2.2. Samples for the comprehensive evaluation

Sample - A - Stock solution II was prepared and diluted with ultrapure water to a final concentration of 500 ng L^{-1}; Sample - B - Stock solution I was prepared and diluted with ultrapure water to a final concentration of 1000 ng L^{-1}; Sample - C - River water sample which was collected from the Danube River at RKI 2568 (24 h composite sample, DOC ≈ 2.7 mg L^{-1}) spiked with stock solution I to reach a final concentration of 100 ng L^{-1}; Sample - D - Stagnant

tap water (DOC ≈ 2.7 mg L^{-1}) collected from a plastic pipe after incubation for 96 h at 40 °C and spiked with stock solution I to obtain a concentration level of 100 ng L^{-1}.

2.2.3. Samples for the final confirmation

A total number of 130 isotopically labeled standards (supplemental table S4) were spiked to four different sample matrices - ultrapure water -E-, groundwater -F-, river water -G- and pretreated wastewater -H-. The ultrapure water was produced by Milli-Q® Advantage A10® system (Merck, Darmstadt, Germany), the groundwater was directly taken from a local well (DOC ≈ 2.0 mg L^{-1}) while the river water sample consisted of a 24 h composite sample which was collected from the Danube River at RKI 2568 (DOC ≈ 2.5 mg L^{-1}). The most complex matrix represents a 24 h composite sample taken after the secondary clarifier of a conventional wastewater treatment plant (725,000 population equivalent, incl. 400,000 from industry, 60,000 m^3 d^{-1} (during sampling), DOC ≈ 9.5 mg L^{-1}). In this particular case, the sample was filtered using a 0.45 μm pore size membrane filter. All four matrices were spiked with isotopically labeled standards at the three concentration levels 25, 100 and 500 ng L^{-1} to cover a relevant magnitude including the 100 ng L^{-1} limit for pesticides according to Council Directive 98/83/EC. Furthermore, the German Federal Environment Agency suggests a health orientation value "for substances that cannot (or can only partially) be toxicologically assessed [...]" of 100 ng L^{-1} [18,19].

To avoid extensive dilution effects, 950 μL of each sample were spiked with 50 μL of stock solution III leading to a dilution of 5% for all matrices. The concentration of the stock solution was varied to reach the desired sample concentration by always adding the same volume.

2.3. LC-MS sample acquisition

All samples were analyzed with a HPLC system (LC20 series Shimadzu, Kyoto, Japan) coupled to the Q-TOF/MS System Triple-TOF™ 5600 (Sciex, Framingham, USA) equipped with a DuoSpray™ Ion Source. Ions were generated by electrospray ionization in positive and negative ionization mode (ESI+/ESI-) in separate runs and monitored within a mass range of *m/z* 100 to 1200. For data-dependent acquisition (DDA), a 250 ms survey scan was acquired followed by a maximum number of 12 individual MS^2 experiments (descending intensity order) covering a mass range of 30–1200 *m/z*. Precursor ions exceeding a threshold of 100 counts per second (cps) that did not appear on a previously generated exclusion list were automatically selected for MS^2. Ions as well as their isotopes were excluded from DDA for a period of 20 s after eight occurrences. The fragmentation was accomplished with a collision energy ramped from 25 to 55 eV within an accumulation time of 64 ms adding to a maximum cycle time of about 1.1 s. The calibrant delivery system enables mass recalibration after every fifth run by infusion of so-called APCI positive (or negative) calibration solution. The column used for the separation was a Zorbax Eclipse Plus C18 (2.1 × 150 mm, 3.5 μm, Agilent, Santa Clara, USA) connected to the precolumn Security Guard AQ C18 (4 × 2 mm, Phenomenex, Aschaffenburg, Germany) which were maintained at a temperature of 40 °C. The flow rate was kept at 0.3 mL min^{-1} with eluents consisting of water (A) and acetonitrile (B) containing 0.1% (v/v) formic acid, respectively. A multi-step gradient with the following parameters was applied: 0 min (2% B), 1 min (2% B), 2 min (20% B), 16.5 min (100% B), 27 min (100% B), 27.1 min (2% B), 37 min (2% B). To ensure a sufficient number of data points over a chromatographic peak, a normal HPLC was preferred as UPLC sacrifices peak width with the consequence that the number of DDA experiments would have to be reduced. For peaks with very low retention

factors, a different chromatographic technique would yield better results. This has, however, not been investigated in this study. The blank value was generated by a zero-injection to cover the system blank value. After averaging all spectra over the complete chromatographic run of the system blank, ions exceeding an intensity threshold of 50 cps were set on an exclusion list to avoid continuous fragmentation of system contaminations throughout the DDA experiments. Note that irrespective of the ionization mode, formic acid was used as eluent additive to ensure equal retention times which is indispensable for a direct comparison between both modes even though the loss of sensitivity in negative mode has to be accepted.

2.4. Software tools

The instrument control software Analyst® TF (1.7) was used for data acquisition based on the described LC-MS method parameters. MarkerView™ (1.2.1) was utilized to perform peak finding as well as alignment of multiple samples while the quantitation package MultiQuant™ (3.0.2) was applied for subsequent peak integration. All mentioned software tools are commercially available from Sciex (Framingham, USA). Other evaluation steps comprising different algorithms were performed with Matlab (R2015a, The MathWorks Inc., Natick, USA).

3. Data processing

3.1. Workflow

In the following, the main steps of the data processing approaches are described. More details and exact parameter settings can be taken from supplemental table S5. The entire strategy to process the data sets presented in this paper is composed as follows:

- I. Peak finding
- II. Peak alignment
- III. Peak integration
- IV. Filtering
  - i. Blank correction
  - ii. Absolute threshold
  - iii. Relative threshold

(I) In the first step, all features (peaks) within a predefined retention time and mass (*m/z*) window exceeding a given noise threshold were set on a peak list. The peak finding was done by an algorithm called *Enhance* [20]. The considered time window is limited by the chromatographic restrictions (e.g. dead time) and the noise threshold was taken from some different spectra throughout the chromatographic run in a way that it was set slightly higher than the noise [20]. For samples analyzed in this study, a noise threshold of 100 cps was generously determined.

(II) After non-target peak finding, the detected peaks representing the same features from different samples were merged - based on retention time and mass tolerances - in the peak alignment whereas no retention time or mass correction has been implemented.

(III) Subsequent peak integration was performed to preserve more peak characteristics such as peak area, peak height and peak width. Software tools well known from target analysis were implemented to integrate all peaks from the raw data that were originally picked by the peak finding algorithm.

(IV) The previous step, however, requires further filter criteria to eliminate false positive findings that were, inter alia, created by the subsequent integration. It should be mentioned that a

non-negligible number of false positives is already produced by the peak finding algorithm which further emphasizes the importance of these measures.

(i.) All features whose peak area and peak height did not substantially differ from the blank's ones were eliminated. The blank - in this study - was stipulated to be a system blank value created by zero-injection. In reality and based on the analytical request, a more convenient choice may considerably simplify the non-target procedure.

(ii.) Furthermore, additional filter criteria which include minimum peak area (A), minimum peak height (H), retention time range (R), range of the full width at half maximum (W) as well as the ratio of peak area and peak height (A/H) were defined on the basis of standard substances (see supplemental table S5). We further assume that unknown compounds, which are ascertainable by the applied method, show similar behavior. These criteria are referred to as absolute threshold since one single injection is sufficient to apply these filters. Features that did not meet the criteria of these absolute thresholds were not taken into account.

(iii.) Finally, the relative thresholds were applied to also consider the variations of the peak characteristics. For instance, the peak areas of real features in several technical replicates are subjected to statistical variations which are not likely to exceed a certain value (device-dependent) while, on the other hand, false positive features which are predominantly created by matrix or noise signals show substantially higher fluctuations. In contrast to real peaks, this circumstance is almost exclusively attributable to unrepeatable integration limits detected for such signals. To also address this issue, several models covering these relative thresholds were evaluated and compared to each other.

Note that the single steps were fully automated whereas the complete workflow could be automated as well.

3.2. System stability

It has to be ensured that the system - comprising LC separation as well as MS detection - permits the comparison over n injections without discriminating against single ones. Problems such as column bleed, carryover, major system fluctuations or loss of sensitivity over time (trend among consecutive injections) would considerably impair this approach. In this study, a number of $n = 21$ technical replicates have been chosen in order to regard features right down to a recognition rate of less than 5%. One of the models requires three different injection volumes (here: 50, 75 and 100 μL) leading to a total number of $n = 63$ injections. Therefore, the stability over 63 injections was proven based on 263 standards (supplemental table S2) at a concentration of 500 ng L^{-1}. This relatively high concentration level was chosen to cover as many substances as possible - even low-response candidates such as parathion-methyl. Splitting the data set according to the three different injection volumes, it could be shown that more than 95% of the standards reveal coefficients of variation (CV) of less than 10% at each of the three injection levels (supplemental figure S1). Even if regarding all 63 injections chronologically (almost 40 h total analysis time) in a way that the sums of the relative peak areas were considered, the deviations with respect to the expected value of $N = 263$ were always less than 3.5% (supplemental figure S1). The same data set of 63 injections was processed with a non-target approach leading to a total number of $N = 7337$ features. To

check whether or to what extent a trend is to be noticed, the trend test according to von Neumann [21] was performed. For this purpose, the data set of 63 injections was divided according to the injection volume resulting in three sub data sets which were regarded for the trend test. Starting with four replicates, which is the minimum sample size to be compared to the test value, the number of considered injections was successively increased until all $n = 21$ replicates were included. For all three injection levels, the maximum proportion of features showing a trend, hence possessing a test value higher than the critical limit (for $p = 99\%$) according to von Neumann, was less than 10%. Moreover, a further data set comprising 63 zero-injections (system-blank) was consecutively analyzed which allows to perform the trend test over all $n = 63$ replicates. Again, starting with four replicates, the critical limit for all $N = 2295$ features was successively calculated and compared to the tabulated values, until all 63 injections were included in the sample. The maximum number of features revealing a statistically significant trend for $p = 99\%$ was less than 6.5% (supplemental figure S2). These findings indicate sufficient system stability with only low susceptibility to the above mentioned problems which allows the evaluation of different models on the basis of the repeatability.

3.3. Processing models

The five different data processing models (a) - (e) used for the comparative evaluation are introduced in more detail. To evaluate the repeatability, a combinatorial approach was used to compare all theoretically possible combinations of a given set of technical replicates. The general formula to calculate the total number of combinations C_{max} for all models used in this study was derived to be:

$$C_{max} = \binom{n}{k}^u \quad \text{with } n \geq k, \ u \geq 1 \ \text{and} \ n, k, u \in \mathbb{N} \tag{1}$$

The total number of combinations C_{max} is dependent on the number of technical replicates n, the number of samples k which are chosen from the set of replicates and the number of different injection volume levels u (e.g. 50, 75 or 100 μL). This procedure is the analog to the well-known urn problem where n is the number of balls, k the number of draws and u the number of urns. The models and the respective combinatorics are schematically shown in Fig. 1.

The different characteristics of the models used for the evaluation are summarized in Table 1. Single injections (a), duplicates (b), triplicates (c), quadruplicates (d) and triplicates at three different injection levels (e) were considered during the data evaluation. The fact that just one single replicate is used implies that model (a) does not consider any relative threshold criteria and hence the variation of the peak characteristics remain disregarded (workflow without IV. iii.). For the remaining models, all possible quotients of the respective parameters (A, H, R, W and A/H) are calculated between all considered replicates (e.g. for model (c): #1/#2, #1/#3 and #2/#3). Features with just one quotient out of range were neglected. The evaluation workflow for model (b) is illustrated in an animation.

For model (e) one replicate ($k = 1$) at each injection level ($u = 3$) is picked to assemble the subsample. In this case, the three quotients are only calculated for parameters affected by the concentration (hence, the injection volume) which are the peak area A and the peak height H. In contrast to the other models, the ratios were calculated by dividing the parameter (A and H) that are multiplied with the reciprocal volumes leading to the equation (2):
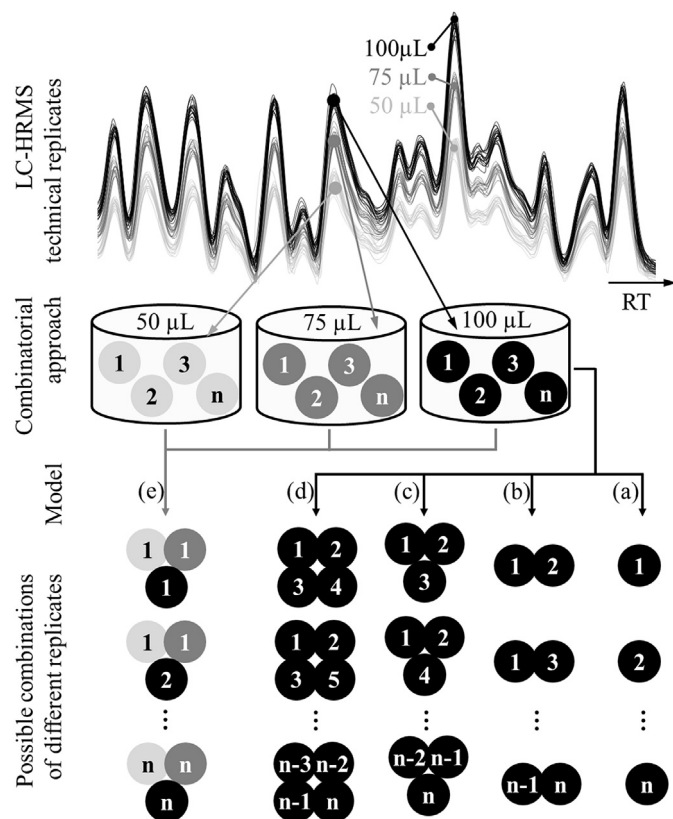
Fig. 1. Exemplary illustration of the models used in this study; *n* = number of technical replicates; (a) single injection (*k* = 1), (b) duplicates (*k* = 2), (c) replicate triplicates (*k* = 3), (d) replicate quadruplicates (*k* = 4) and (e) replicates are chosen from three different injection volumes (*u* = 3, *k* = 1); Note: for model (e), the total number of injections equals *n·u*.

$$Q_{n,k} = \frac{X_n}{X_k} \frac{V_k}{V_n} \qquad (2)$$

while X_i either represents the peak area (*A*) or the peak height (*H*) and V_i the injection volume at the respective level. The threshold values for each $Q_{n,k}$ were derived by assuming that the maximum deviation of the $(n-1)$-th level must not exceed the minimum deviation of the *n*-th level. Thus, for each of the three quotients extreme values were computed and features that did not meet those criteria excluded from further steps. It should be noted that the number of replicates n in this particular case represents three individual injections with different injection volumes (*u* = 3) adding to a total number of *n·u* = 63 runs.

For each single combination of various replicates, the complete

workflow (blank subtraction, absolute and relative thresholding) was performed using Matlab.

Converting the number of combinations (C) to the number of remaining replicates (r)

Due to the fact that each model follows another combinatorics, a direct comparison of the combinations among the several models is not possible. In contrast to the number of combinations (which differ between the models, see Table 1) the number of replicates is common to all which is why the comparison has to be conducted at this level. However, in reality some features do only partly fulfill the different threshold criteria, leading to combination values *C* which are smaller than C_{max}. Consequently, the number of all replicates (*n*) is reduced to the number of the remaining replicates $r = n - x$ with *x* being the number of replicates for which the feature did not pass the thresholds. From the combinatorial perspective, a function $f(r) = C$ can be derived for model (a) to (d) which maps each number of remaining injections (*r*) to the corresponding combination value (*C*). The inverse function $f^{-1}(C) = r$, on the other hand, allows to map each combination value to the number of remaining replicates. For Model (b), the animation exemplifies the complete processing including this conversion step which is explained in more detail in the supplemental explanation S1. Model (e), however, demonstrates a special case as, depending on the distribution of the replicates that do not pass the thresholds to the three injection levels, multiple numbers of combinations are obtained for the same number of remaining replicates (*r*). This circumstance is discussed extensively in the supplemental material (explanation S1). Nevertheless, to enable the comparison to other models, in this particular case, the best- and worst-case scenarios for the remaining number of injections were calculated and taken into consideration.

3.4. Characteristics derived for the comprehensive evaluation

Compared to target analysis, the general problem in the non-target approach is that we do not know the ground truth about how many signals (at which *m/z* and retention time) are to be expected. The lack of an appropriate point of reference which should be achieved makes it difficult to assess common validation characteristics. For instance, based on a non-target approach it is not possible to conventionally report the false negative rate while extensive effort would have to be invested to manually search for false positives. Instead, the strengths and weaknesses of different processing strategies were evaluated by comparing the repeatability over *n* technical replicates. To assure that features of interest were not removed by the defined filtering criteria, a chosen set of substances was spiked to the samples examined in this work. It should thus be possible to at least partly evaluate the false negative rate based on some target compounds. Later, the evaluation of partially false negative findings using various technical replicates

Table 1
Summary of the different processing models used for comparative evaluation.

	Processing models				
	(a)	(b)	(c)	(d)	(e)
Number of considered replicates (*k*)	1	2	3	4	1
Number of different injection levels (*u*)	1	1	1	1	3
Injection volumes [µL]	100	100	100	100	50, 75, 100
Blank correction	✓	✓	✓	✓	✓
Absolute thresholding	✓	✓	✓	✓	✓
Relative thresholding	✗	✓	✓	✓	✓
Number of quotients for relative thresholding	✗	1	3	6	3
C_{max} (for *n* = 21)	21	210	1330	5989	9261

will be introduced. Note that deisotoping and adduct assignment was not performed in any of the presented workflows, as these signals also contribute to the repeatability. Furthermore, the processed data were compared against the raw data to ascertain whether and to what extent improvement of the data quality appears.

Prior to the comparative evaluation, meaningful characteristics were defined which allow assessing the data on the basis of few parameters. The mean rate of recognition as well as the mean improvement factor proved to be informative values to characterize the different models of data processing. The derivation of these two values is explained below.

3.4.1. Repeatability, mean rate of recognition ($\overline{RR}$)

In order to evaluate the repeatability, the rate of recognition (RR), which describes how often one particular feature occurs over n replicates was defined. After peak alignment, the data are arranged in a matrix containing m features (rows) and n samples (columns) while the elements $A_{i,j}$ represent the respective response.

$$\underline{D} = \begin{pmatrix} A_{1,1} & \cdots & A_{1,j} & \cdots & A_{1,n} \\ \vdots & & \vdots & & \vdots \\ A_{i,1} & \cdots & A_{i,j} & \cdots & A_{i,n} \\ \vdots & & \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,j} & \cdots & A_{m,n} \end{pmatrix} \tag{3}$$

To assess the feature distribution over n technical replicates, the rate of recognition (RR_i) was calculated for each feature. In a first step, all $A_{i,j}$ that did not exceed predefined thresholds were set to zero. The general formula can therefore be written:

$$RR_i = \frac{1}{n} \sum_{j=1}^{n} sgn(A_{i,j}) \tag{4}$$

Applied to all features, a mean rate of recognition ($\overline{RR}$) is derived:

$$\overline{RR} = \frac{1}{m-z} \sum_{i=1}^{m} RR_i = \frac{1}{(m-z)\cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} sgn(A_{i,j}) \tag{5}$$

where z represents the number of lines i that only contain zeros (which may appear if certain filter criteria are applied). To enhance the comparability between different data sets, the feature distributions are shown as cumulative stairstep graphs.

Ideally, each individual feature would exactly occur n times ($\overline{RR} = 1$), however, this circumstance might significantly differ due to various difficulties in peak recognition and/or in measurement. One of the most important parameter affecting the $\overline{RR}$ is the intensity threshold to distinguish between noise and real peaks. Setting a low intensity threshold decreases the repeatability as many noise signals are recognized as peaks (false positives), whereas using a high intensity threshold produces better repeatability, however, real peaks of interest might be missed (false negatives). A typical feature distribution over 21 consecutive injections of a water sample is illustrated in Fig. 2 where four different intensity thresholds were used to process the same data set. Note that in order to demonstrate the difficulties, only the first two steps of the entire workflow were performed. The distribution processed with an intensity threshold of 100 cps clearly reveals the deviation from the ideal. On the other hand, the gain in repeatability using higher thresholds can clearly be recognized.

This relationship clearly demonstrates the substantial influence of this parameter on the feature distribution. However, even though the $\overline{RR}$ significantly increases (Table 2), the total number of

recognized peaks, on the other hand, is considerably reduced and might cause the loss of less abundant species.

It seems obvious that this contrary behavior requires a compromise between the repeatability and the total number of features recognized. However, the authors believe that the lack of repeatability — especially for low intensity thresholds — might be bypassed by improving the data processing methods after the original peak finding (workflow step III and IV).

3.4.2. Mean improvement factor ($\overline{I}$)

Aside from the mean rate of recognition which characterizes the feature distribution, the general improvement with respect to the raw data should also be taken into consideration. In order to accomplish this, the occurrence of the features in n replicates was regarded before and after data processing. To calculate the mean improvement factor ($\overline{I}$), for each single feature the Δ recognition value (= times recognized after processing - before processing) was formed and sorted in descending order (Fig. 3).

Data points above the zero line (Fig. 3 (b), dashed line) demonstrate features with increasing recognitions throughout the processing whereas the ones beneath indicate deterioration. Data points lying on the zero line itself are not subjected to any changes which predominantly consist of features occurring in all replicates before as well as after the processing. Instead of regarding the mere number of features showing improvement or deterioration, the extent of the changes should be associated as well.

To include this fact, the areas under the curve reveal the number weighted by the extent of improvement or deterioration ($Area(+)$ for improvement, $Area(-)$ for deterioration). The mean improvement factor comprises as follows:

$$\overline{I} = \frac{|Area(+)| - |Area(-)|}{|Area(+)| + |Area(-)|} \tag{6}$$

Features contributing to $Area(+)$, hence peaks which were more frequently recognized after processing represent partially false negative findings since the peak finding algorithm at least once missed a real peak. This offers the opportunity to recognize false negatives findings on the basis of non-target features as an interesting alternative to the spiked compounds.

4. Results and discussion

4.1. Comprehensive evaluation

For the comprehensive evaluation of the models, four different samples (-A- to -D-) were used (see chapter 2.2.2). To cover different matrix compositions, river water and stagnated tap water (plastic pipe) were chosen in addition to aqueous standard solutions. Each sample was analyzed 21 times with three different injection volumes (50, 75 and 100 μL) hence leading to 63 injections. For models (a) - (d), only the 21 injections with an injection volume of 100 μL were regarded whereas model (e) comprises all 63 injections. After running through the models, the output represents a list containing all features recognized by the algorithm as well as a counting number within 0, i.e. feature rejected in each single cycle, and the number of all theoretical combinations C_{max}, i.e. feature accepted in each single cycle, meaning that the peak characteristics (A, H, W, R and A/H) as well as their relative fluctuations have fulfilled all specified filter criteria. The integer counting number (C) can be converted - as has been described - to the number of remaining injections (r). Further evaluations and comparisons are based on these converted numbers of remaining injections. Features which represent spiked standards ($m/z \pm 10$ ppm, $RT \pm 0.15$ min) were extracted from the result matrix and considered
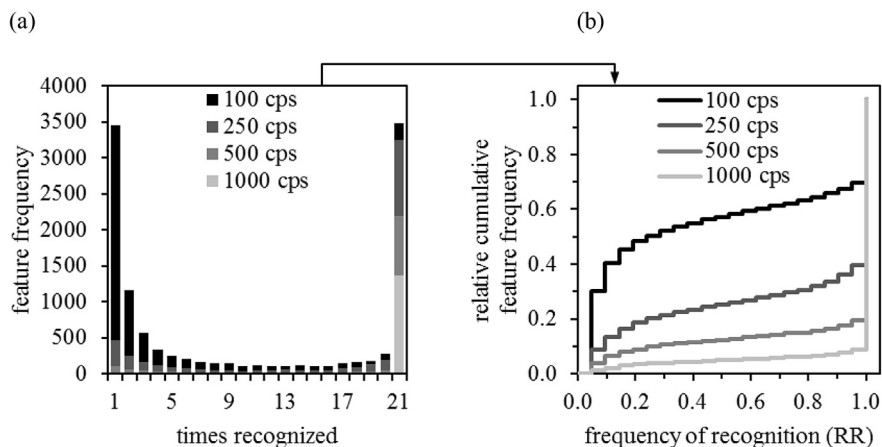
Fig. 2. Feature distributions as histograms (a) and as cumulative stairstep plots (b) over $n = 21$ technical replicates of a water sample processed with four different intensity thresholds (100, 250, 500 and 1000 cps); only peak finding and peak alignment were performed.

Table 2
Influence of the intensity threshold on the total number of features detected and $\overline{RR}$ ($n = 21$).

Intensity threshold [cps]	Total number of features	Mean rate of recognition ($\overline{RR}$)
100	11,450	0.467
250	5387	0.760
500	2733	0.882
1000	1501	0.952

Note: Only peak finding and peak alignment were performed.
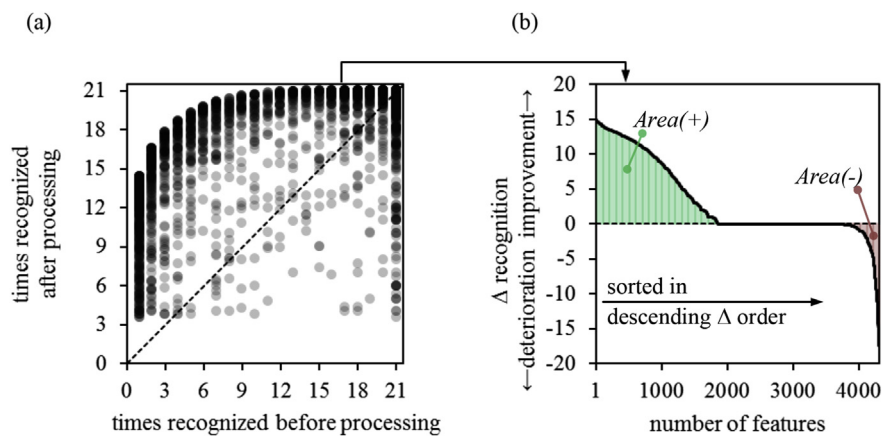
Fig. 3. Comparison of features before and after processing.

separately. The number of spiked standards which are eliminated by the threshold criteria was expected to be very low as these filters have been designed taking into account a large data set of known standards (data not shown). These standards were selected to be compatible with ESI as well as RPLC, particularly with regard to cover a wide range of polarity (logP $\approx$ −1 to 5). Furthermore, the $\overline{RR}$ of the spiked standards was estimated to almost reach 100%. In addition to the numbers obtained by the models, the raw data were regarded as well, since features which were not even recognized once do not contribute to the list used for processing for which reason the number of features recognized in the raw data has to be regarded as the maximum possible one. The same evaluations were performed for all non-target features. The remaining number of features and the corresponding $\overline{RR}$ were determined while the comparison to the unprocessed raw data reveals the mean improvement factor. The results of the entire evaluation are tabulated in Table 3.

4.1.1. Discussion (comprehensive evaluation)
The overall recognition of the spiked standards turned out to be very good since no more than two of the spiked standards were missed by the peak finding algorithm. A closer look revealed that in case of stock solution I, Etofibrate (sample -C- and -D-) and Iopamidol (sample -D-) were not recognized due to signal suppression (matrix effects). Regarding stock solution II, Irgarol-descyclopropyl (Metabolite M1) as well as Desethylterbutylazin were not recorded even though clear signals were manually determined. The recognition in both cases failed due to background subtraction which was accomplished by regarding the MS spectra 10 scans before the actual peak apex. However, on both ion traces, a structural isomer (Desmetryn and Simazin, respectively) elutes about 10 scans before the mentioned standards hence leading to elimination of the latter

Table 3
Comprehensive evaluation of four different data sets.

Sample	-A-	-B-	-C-	-D-
Matrix	Ultrapure water	Ultrapure water	River water	Stagnated tap water
Concentration [ng L^{-1}] of spiked standards	500	1000	100	100
Stock solution #	II	I	I	I
# of standards	263	32	32	32
a. Target compounds (spiked)				
Number of standards				
Raw data (a–d)	261	32	31	30
Raw data (e)	261	32	31	30
Model (a)	261	32	31	30
Model (b)	261	32	31	30
Model (c)	260	32	31	30
Model (d)	260	32	31	30
Model (e)	261	32	31	30
Mean rate of recognition ($\overline{RR}$)				
Raw data (a–d)	0.996	1.000	0.988	0.998
Raw data (e)	0.993	1.000	0.954	0.965
Model (a)	0.988	1.000	0.986	0.967
Model (b)	0.985	0.985	0.995	0.964
Model (c)	0.987	0.982	0.994	0.961
Model (d)	0.987	0.980	0.994	0.960
Model (e)	0.979–0.984	0.977–0.984	0.975–0.978	0.948–0.962
b. Non-target compounds				
Number of features				
Raw data (a–d)	6672	4773	2372	5088
Raw data (e)	7337	5551	2655	5582
Model (a)	4441	2135	844	2978
Model (b)	4369	1891	815	2899
Model (c)	4298	1841	747	2822
Model (d)	4215	1688	667	2725
Model (e)	4301	1791	743	2774
Mean rate of recognition ($\overline{RR}$)				
Raw data (a–d)	0.637	0.568	0.596	0.592
Raw data (e)	0.517	0.448	0.484	0.471
Model (a)	0.694	0.541	0.533	0.605
Model (b)	0.797	0.671	0.639	0.725
Model (c)	0.840	0.708	0.714	0.775
Model (d)	0.866	0.767	0.792	0.809
Model (e)	0.739–0.812	0.612–0.719	0.581–0.682	0.658–0.757
Mean improvement factor ($\overline{I}$)				
Model (a)	−1.000[a]	−1.000[a]	−1.000[a]	−1.000[a]
Model (b)	0.668	0.319	0.412	0.598
Model (c)	0.724	0.409	0.583	0.657
Model (d)	0.738	0.470	0.661	0.671
Model (e)	0.661–0.826	0.437–0.759	0.641–0.830	0.636–0.829

[a] By definition $-100\% \left(\lim_{|A(+)| \to 0} \overline{I} = \frac{|A(+)| - |A(-)|}{|A(+)| + |A(-)|} = -1 \right)$.

ones and, consequently, to false negative findings (see supplemental figure S6).

As the data basis differs between models (a) - (d) (21 injections) and model (e) (63 injections), different raw data must be taken as a reference (see Table 3). Considering the different models, it is conspicuous that almost all standards (>99.6%) were retained, which underlines the general applicability of all presented models and thresholds involved. Even for the more complex matrices in case of sample -C- and -D-, concentrations of 100 ng L^{-1} were sufficient to confirm all recorded standards by each model. In most instances, the health orientation value of 100 ng L^{-1} as suggested by the German Federal Environment Agency [18,19] is not impaired by any of the filter criteria.

Moreover, the mean rate of recognition for the target compounds shows that the applied models have little or no effect on the repeatability of the method. Also the $\overline{RR}$s derived from the raw data delivers outstandingly good values. The better-than-expected $\overline{RR}$s obtained for the raw data sets suggest that - in general and particularly for real peaks with sufficient abundance - the peak

finding algorithm works consistently. These findings clearly demonstrate that features of interest were barely or not removed by the filters which suggest that the false negative rate, i.e. real peaks which were eliminated by the filters, is not significantly affected by the processing strategies.

While the $\overline{RR}$s for target compounds were not showing any weaknesses and thus no need of extensive processing strategies, this circumstance differs significantly if regarding non-target features where, on average, a deterioration of more than 39% (model (a) - (d)) and more than 49% (model (e)) could be observed when comparing the raw data sets (target vs. non-target compounds). These data clearly point out the need for action if processing non-target data at trace level. The effect of the derived filter criteria on the total number of features is distinctly identifiable when comparing the remaining numbers of features with the ones tabulated for the associated raw data. Note that the numbers of recognized features for the raw data of model (e) always reveal higher values that are caused by the 42 additional injections (50 and 75 μL). Taking into account models (a) - (d), on average, 35.1%

and 43.9% of the features were removed from sample -A- and -D-, respectively while 60.4% and 67.6% of the features failed the specified threshold criteria if regarding sample -B- and -C-. Since model (c) and model (e) both comprise three replicates in the subsamples, we expected the number of features to be in a very similar range. While the numbers indeed resemble each other, it is surprising that, except for sample -A-, the remaining numbers obtained with model (e) are slightly lower than the ones obtained with model (c) which is much likely due to the lower effective concentration caused by the smaller injection volumes. It should also be noted that the number of detected features does not necessarily permit conclusions regarding the complexity of the sample as it was deliberately refrained to perform deisotoping and adduct assignment, as these signals also contribute to the repeatability. This fact, however, increases the number of features recognized but not inevitably to the same extent between different samples. For instance, the relatively high number of 6672 features initially detected in sample -A- (raw data (a)-(d)) impressively illustrates the increase in the number of features (only 263 substances spiked). The much higher number may possibly be explained by isotopes, adducts, dimers, in source fragmentation or multiple charge states. Furthermore, accompanying substances or contaminations in the standard solutions or in the ultrapure water, which was used to dilute the standards, may also contribute to the higher number of features. It is therefore not surprising that the number of features detected in the river water is smaller than the ones in the aqueous standard solutions which contain much higher concentrations and hence increase the probability that such species exceed a given intensity threshold.

To further assess the different data sets, the mean rate of recognition $\overline{RR}$ was taken into consideration. The evaluation of the non-target compounds is exemplary demonstrated using sample -A- (Fig. 4), the diagrams of the remaining samples are illustrated in supplemental figures S7-9. As a first step, the $\overline{RR}$ was derived for the raw data and for each single model. It should be noted that due to the conversion (C to r) also non-integer numbers for the remaining injections were obtained, thus leading to a quasi-continuous distribution in contrast to the discrete ones obtained for the raw data and model (a). As discussed before, the best- and worst-case scenarios were assigned for the evaluation of model (e).

The flatter courses of the cumulative distributions clearly reveal the improvement compared to the raw data as is also evident from the $\overline{RR}$s. The fact that the rates increase for model (a) to (d) is not surprising as more replicates per subsample were considered during the evaluation. The poorer results obtained for model (e), even if considering the best case scenarios, are much likely attributable to the three-fold increase in the number of injections and hence the possibility of generation of false positive findings. Nevertheless, notable improvements can be identified if comparing the distribution of model (e) with the raw data. Regarding all samples, for model (e) mean improvements of the $\overline{RR}$s of almost 35% in the worst-case scenario and 55% in the best-case scenario with respect to the raw data were perceived. Assessing model (b) - (d), the comparison to the unprocessed raw data clearly points to an improvement in repeatability. Based on the raw data, a relative increase in the $\overline{RR}$s, averaged over the four sample types, of more than 18%, 26% and 35% could be demonstrated for model (b), (c) and (d), respectively. Model (a), on the other hand, only shows minor improvement for sample -A- and -D- whereas the rates for -B- and -C- even deteriorate. This effect is much likely attributed to false positive findings in the raw data which result in apparently better $\overline{RR}$s. Such candidates were sometimes incompletely eliminated during the absolute thresholding and can therefore, should the situation arise, even result in lower $\overline{RR}$s.

As can be seen, however, the gradual increase in the $\overline{RR}$s becomes lower the more technical replicates were used. While model (b) offers - with respect to model (a) - a mean improvement of the $\overline{RR}$s of more than 11%, only 5%, on average, were gained among model (b) → (c) and (c) → (d), respectively. The high increase between model (a) and (b) clearly indicates the strength of the relative threshold criteria (i.e. the fluctuation of the peak characteristics which are regarded in models (b) to (d)) whereas more replicates increase the $\overline{RR}$s to a smaller extent. These results demonstrate that the initial problems were not primarily caused by the measurement but rather by the problems in peak recognition. The comprehensive data processing strategy, however, partially compensates the appearing difficulties using only few replicates. The evaluation on the basis of the $\overline{RR}$ combines both, the elimination of false positives as well as the correction of features which were only partly recognized. It remains to be seen whether the improvement in $\overline{RR}$ is mainly caused by the first or the second option. The mean improvement factor, on the other hand, accounts for the latter ones and is described in the following.

The occurrence of the features in n replicates was regarded before and after data processing. This comparison reveals the extent of the improvement or deterioration for the different models. For sample -A-, the evaluation is depicted in Fig. 5 and described in more detail below while the evaluation of the
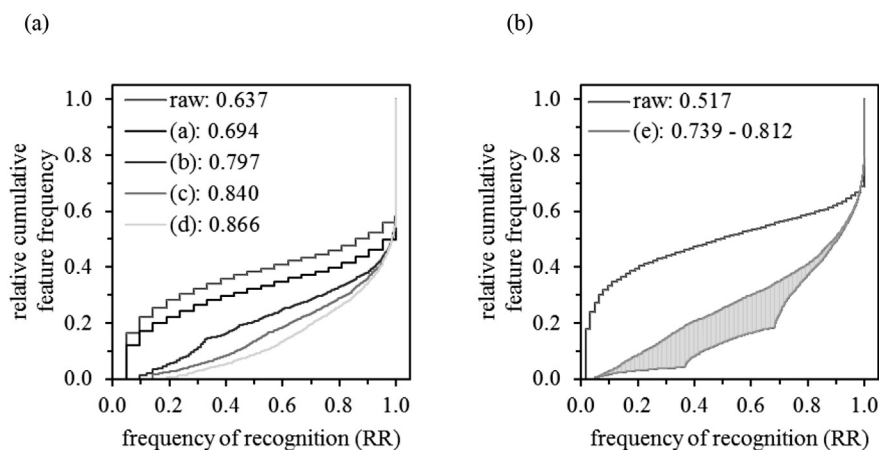
Fig. 4. Cumulative feature distributions obtained for sample -A-, (a) distribution of unprocessed raw data (21 injections, 100 μL) and for models (a)–(d), (b) distribution of unprocessed raw data (63 injections, 50, 75, 100 μL) as well as for model (e); $\overline{RR}$s in the legend are depicted after the respective entry.

remaining ones can be taken from supplemental figures S7-9. The fact that the intersect increases for model (a) - (d) results from the combinatorial approach. Once a feature has been recognized, the subsequent peak integration, in the best case, is always leading to positive results for all combinations including this particular replicate. The maximum Δ recognition value can also be derived from the combinatorial approach and increases the more replicates are taken into consideration. While model (a), by definition, never shows an improvement, maximum Δ recognition values of about 9.0, 12.8, 14.8 and 42.0 could theoretically be reached using model (b), (c), (d) and (e), respectively (see supplemental figure S10). These values calculated from a theoretical approach can also be found if processing real data (see intersects Fig. 5).

The mean improvement factors - with respect to the raw data - are summarized in Table 3. Regarding the tabulated factors it is obvious that for model (b) - (e), irrespective of the sample type, a clear improvement with respect to the raw data could be achieved.

The evaluation was performed in a way that only features which have been recognized at least once (raw data) in each selected subsample were further taken into consideration. It is thus evident that model (a), comprising only single replicates, will never lead to an improvement which is why the mean "improvement", by definition, will always report - 100%. For model (a), however, the number of features subjected to deterioration was always less than 15% for all sample types (see supplemental table S6). These are much likely features which were partly removed by the filter (no relative threshold) resulting in negative Δ recognition values. For model (b) to (e), on the other hand, the $\bar{I}$s reach positive values between 31.9% and 73.8% (83.0% in the best-case scenario) which, once again, considerably demonstrate the benefits when regarding multiple replicates (see Table 3). The total numbers of features showing improvement, deterioration or no change without being weighted by the extent are tabulated in supplemental table S6 which also comprises the areas used to calculate the $\bar{I}$s.

In summary, different aspects have to be considered: while the recoveries as well as the $\overline{RR}$s of all spiked standards clearly show the outstandingly good coverage and thus no need of action, the picture changes dramatically if regarding the non-target compounds. The recognition frequency shows consistently worse values but could successfully be increased by application of the different processing models. On the other hand, the major problem of partially false negative findings becomes apparent only after considering multiple replicates. Against the background that the recovery of spiked standard almost reaches 100%, it was all the more unexpected that partially false negative findings occurred to this extent.

Nevertheless, the mean improvement factors clearly illustrated the advantages if considering multiple replicates. Moreover, the number of retained features was significantly reduced during the processing which simplifies the non-target screening. The relative threshold criteria, i.e. the fluctuation of the peak characteristics as well as the consideration of multiple replicates have emerged to be very powerful when dealing with type I and type II errors. While the recognition frequencies as well as the improvement factors increase the more replicates were considered, the benefits must be balanced against the decrease in sample throughput. Based on these findings, it is strongly recommended to use at least two, or better, three technical replicates while further consideration should be given to the relative threshold criteria. We found that model (c) reflects an acceptable compromise between the benefits ($\overline{RR}$ and $\bar{I}$) and the sample throughput.

Note that real sample matrices were spiked with standard solutions at a certain concentration level. However, it may be true that some compounds were already present in the untreated samples, thus leading to higher effective concentrations than tabulated. For this reason, isotopically labeled standards were used for the final confirmation experiment, which allows to precisely regard the concentration.

4.2. Final confirmation

A total number of 130 isotopically labeled standards (IS) was spiked to four different matrices - ultrapure water (-E-), groundwater (-F-), river water (-G-) and secondary clarifier of a conventional wastewater treatment plant (-H-) - at three concentration levels - 25, 100 and 500 ng L^{-1}. The manual evaluation at the highest concentration level revealed that 123 standards could be detected in ESI positive mode while 56 were recognized in the negative ionization mode (48 in both modes). The isotopically labeled standards with their corresponding retention time are tabulated in supplemental table S4. As expected, the intensities of the total ion chromatograms increase substantially with matrix complexity (Fig. 6). These findings correlate with the concentration of the dissolved organic carbon in the samples. The sample collected from the secondary clarifier -H- clearly represents the most complex one for which reason we expected to have the strongest matrix effects and hence, the greatest difficulties during peak detection of the isotopically labeled standards.

It was found that many substances (e.g. Fig. 6 b, c) are only subjected to minor fluctuations and hardly reveal any influence of the sample matrix. While some of the EICs follow an almost

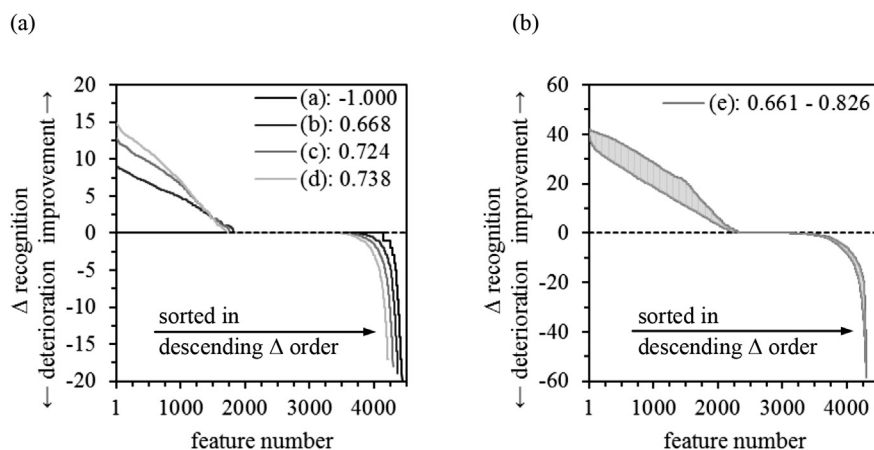(a)   (b)

Fig. 5. Comparison of the feature recognition before and after data processing (Δ recognition) for (a) models (a)–(d) and (b) model (e); $\bar{I}$s in the legend are depicted after the respective entry.

identical course over all matrices, on the other hand, some candidates demonstrate problems when dealing with complex matrices. Either poor selectivity (e.g. Fig. 6d) or strong signal suppression (e.g. Fig. 6e) complicated the assessment of the obtained data and thus resulted in a lower number of recognized standards. To finally confirm the presented models, each matrix at each concentration level was analyzed six times (50, 75 and four times 100 μL injection volume) to obtain the minimum required number of injections to apply each model, leading to a total number of 72 injections for each ionization mode. The parameters used to process the data were maintained as stated previously (supplemental table S5). The obtained results are summarized in Table 4, where the recorded numbers of standards N_s are tabulated for the different models, in four different matrices and for three concentration levels. Moreover, part a. comprises data recorded in the positive ionization mode whereas the results obtained in the negative one can be found in section b. The complete evaluation is available in the supplementary material (SM_2.xlsx) where the results are illustrated for each substance. The individual entries were created as follows:

Manual peak finding

A suspect target method was established using the exact mass ($m/z \pm 10 ppm$) as well as the expected retention time ($\pm 0.15 min$) of each spiked reference standard. The retention time and mass tolerance was evaluated in a former study revealing that the maximum deviation of the mass and the retention time over a time period of 10 months was less than 10 ppm and 0.15 min, respectively (data not shown). After processing the raw data, each single standard (i.e. each EIC) in each matrix and concentration level was manually verified. Therefore a signal-to-noise ratio of about three was considered while only taking into account these standards

exceeding a peak height of 100 cps. At this point it is important to mention, that other filter criteria (e.g. peak width or fluctuation of the peak area) were not taken into consideration. The results obtained during this step were later used to calculate the number of false positives (i.e. substances could manually not be verified while further processing steps indicate them to be real hits). For the final confirmation, the number of false positives, inter alia, is an important factor which previously could not be considered when comparing the different models as the manual peak verification would have been too time consuming. The tabulated values are the sum of all standards detected (under the specified conditions) in the respective sample and concentration level. In this case, the weighted numbers over the four conducted 100 μL - injections are listed (i.e. a standard which was only detected in 3 out of 4 replicates, will be counted as 0.75).

Non-target peak finding

Automated non-target peak finding was performed for all samples whereas features representing the isotopically labeled standards were extracted from the result table (all other features were not considered at this point). The elimination of false positives, by comparing the obtained data to the manually verified ones was performed subsequently. The actual number of recognized standards is tabulated followed by the number of false positives given in brackets behind. Again, the numbers of the recorded features as well as the numbers of false positive findings represent the weighted numbers calculated from the four 100 μL - injections. These data were later used for the correction of other data sets. This is strictly necessary due to the batch-processing, i.e. feature lists are created overall various samples. The objective was to illustrate the data in the way it would have been if only the respective set of samples were processed.
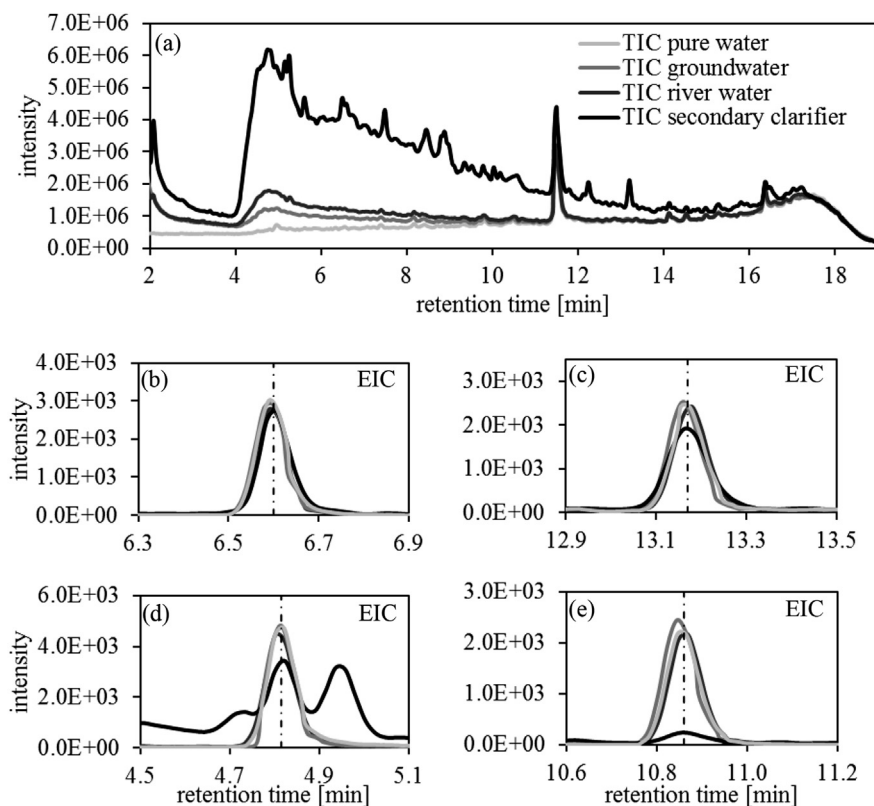
Fig. 6. (a) Total ion chromatograms (TIC) of the different matrices spiked with 100 ng L^{-1} of 130 isotopically labeled internal standards (only a single 100 μL - injection is shown) in positive ionization mode within the evaluated time domain; extracted ion chromatograms (EIC) of Desethylatrazine (b), Mefenamic acid (c), Carbendazim (d) and Valsartan (e), 100 ng L^{-1} each; the dashed lines indicate the expected retention times.

Table 4
Final confirmation with isotopically labeled standards (N_{S+} = 123 in ESI+ and N_{S-} = 56 in ESI-) in four different matrices (ultrapure water, groundwater, river water and pretreated wastewater) and at three different concentration levels (25, 100 and 500 ng L^{-1}).

Number of standards N_{S+} recorded by the respective procedure (false positives)												
Sample	-E-			-F-			-G-			-H-		
Matrix	Ultrapure water			Groundwater			River water			Pretreated wastewater[a]		
Concentration [ng L^{-1}]	25	100	500	25	100	500	25	100	500	25	100	500
a. ESI positive (N_{S+} = 123)												
Manual peak finding[b,c]	114	119	122	115	120	123	114	118	122	89	112	121
Non-target peak finding[b]	95 (0)	114 (0)	121 (0)	99 (0)	116 (0)	121 (0)	93 (1)	115 (0)	120 (0)	67 (6)	104 (2)	118 (0)
Model (a) – manual	96	115	122	99	117	121	92	115	120	67	105	118
Model (a) – automated	93 (0)	111 (0)	121 (0)	98 (0)	114 (0)	121 (0)	91 (0)	115 (0)	120 (0)	66 (6)	104 (1)	117 (0)
Model (b) – manual	97	115	122	104	117	122	95	115	121	70	107	118
Model (b) – automated	92 (0)	111 (0)	121 (0)	99 (0)	114 (0)	121 (0)	93 (0)	115 (0)	121 (0)	63 (4)	101 (1)	116 (0)
Model (c) – manual	101	115	122	104	117	122	97	115	121	74	107	120
Model (c) – automated	92 (0)	111 (0)	121 (0)	96 (0)	114 (0)	121 (0)	92 (1)	115 (0)	121 (0)	60 (3)	99 (1)	116 (0)
Model (d) – manual	101	115	122	104	117	122	100	116	121	75	107	120
Model (d) – automated	91 (0)	111 (0)	121 (0)	96 (0)	114 (0)	121 (0)	93 (1)	114 (0)	121 (0)	59 (2)	99 (1)	116 (0)
Model (e) – manual	92	115	122	99	117	122	93	115	120	69	108	119
Model (e) – automated	75 (2)	106 (0)	119 (0)	86 (0)	112 (0)	117 (0)	81 (0)	111 (0)	119 (0)	55 (4)	94 (1)	106 (0)
b. ESI negative- (N_{S-} = 56)												
Manual peak finding[b,c]	28	44	56	22	43	55	20	40	52	16	27	44
Non-target peak finding[b]	14 (0)	32 (0)	52 (0)	13 (0)	29 (0)	49 (0)	13 (0)	25 (0)	47 (0)	7 (0)	18 (0)	38 (0)
Model (a) – manual	13	31	52	12	30	48	15	26	47	6	18	39
Model (a) – automated	12 (0)	30 (0)	48 (0)	12 (0)	26 (0)	45 (0)	13 (0)	25 (0)	47 (0)	6 (0)	18 (0)	39 (0)
Model (b) – manual	14	33	53	13	30	49	15	27	47	7	18	40
Model (b) – automated	12 (0)	28 (0)	48 (0)	12 (0)	25 (0)	45 (0)	13 (0)	24 (0)	47 (0)	7 (0)	18 (0)	40 (0)
Model (c) – manual	15	34	53	13	31	49	15	28	47	8	19	40
Model (c) – automated	12 (0)	28 (0)	47 (0)	12 (0)	26 (0)	45 (0)	11 (0)	24 (0)	46 (0)	8 (0)	18 (0)	39 (0)
Model (d) – manual	15	35	53	13	33	49	15	28	47	10	19	40
Model (d) – automated	12 (0)	29 (0)	46 (0)	12 (0)	28 (0)	45 (0)	11 (0)	24 (0)	46 (0)	9 (0)	17 (0)	39 (0)
Model (e) – manual	13	32	52	14	30	48	14	27	47	6	18	40
Model (e) – automated	11 (0)	23 (0)	44 (0)	10 (0)	23 (0)	44 (0)	11 (0)	22 (0)	43 (0)	6 (0)	16 (0)	30 (0)

[a] Collected after the secondary clarifier of a conventional wastewater treatment plant.
[b] Weighted number over the four 100 µL − injections.
[c] Only peaks exceeding a height of 100 cps were regarded.

Model (x) - manual

The manual verification was performed to indicate the combinatorial strength of the models without the use of any filters meaning that the presence of a feature was the only criteria that was considered. For this reason, these values should be seen as the highest achievable recovery with the respective model. The tabulated values are composed as follows: According to the respective model, the data from the corresponding injections were extracted from the data set which was manually verified. Again, to only consider features that would originally have been recorded, the ones that did not appear on the non-target peak list, for the particular selection of samples, were eliminated. The sole difference at this point is that the single occurrence of a certain feature would already be enough to be maintained (e.g. detected in 1 out of 3) which means that the different replicates extracted from the non-target data were OR linked which generally leads to a higher coverage, the more replicates were considered. This is the reason why the tabulated numbers increase (or at least stay equal) from model (a) - manual to model (d) - manual. Model (e), however, is based on other injections (50, 75 and 100 µL) which is why these values are likely to be smaller.

Model (x) automated

The actual data that were processed using all threshold criteria (supplemental table S5) are summarized in these entries. In any case, the numbers are smaller or equal to the ones obtained for model (x) - manual, because aside from the mere presence, further thresholds must be satisfied. Instead of taking the exact masses (calculated ones) of the spiked standards, the accurate masses (measured ones), which were assigned by the peak finding algorithm, were used for further processing steps. As supplied before, the features that have initially not been recorded were excluded while the comparison with those manually verified revealed the number of false positives.

4.2.1. Discussion (final confirmation)

The object of the final confirmation was to illustrate - with respect to the matrix compositions as well as to the concentration - that the applied models barely eliminate the spiked compounds. While the overall recovery rate (i.e. with respect to the total number of standards spiked) is important to assess the applicability of the method, this was not of central interest at this point. Please note that in cases where a spiked standard was not even detected (e.g. due to detection limits or matrix interferences) this cannot be attributed to weaknesses of the applied models.

Positive ionization mode

The manual evaluation revealed that at the highest concentration level of 500 ng L^{-1} at least 121 of 123 spiked substances were detected. Interestingly, the set point of N_{S+} = 123 is reached in the groundwater matrix -F- as a certain ionic strength obviously favors the chromatographic separation thus resulting in better peak shapes (e.g. azithromycin was not detectable in ultrapure water). The pretreated wastewater -H-, on the other hand, emphasizes the problem of matrix effects which, in this case, are either attributed to signal suppression or insufficient selectivity (as stated in Fig. 6) and consequently result in a lower number of detected standards. Nevertheless, at a concentration of 100 ng L^{-1} more than 91% (112 out of 123) of the standards could manually be confirmed. If the wastewater matrix is disregarded, mean recovery rates of 93%, 97%

and 99% were achieved for 25, 100 and 500 ng L^{-1}, respectively. These numbers are to be regarded as maximum attainable since each single peak was manually reviewed against the defined requirements. The difference to the total number of spiked standard is attributed to detection limits as well as matrix effects.

The automated non-target peak finding shows similar trends, albeit somewhat lower. The relative mean differences to the manual peak finding, particularly for higher concentrations, are with about 4% and 2% for 100 and 500 ng L^{-1}, respectively, almost negligible. For 25 ng L^{-1}, on the other hand, the mean difference of nearly 16% reveals a significant deviation. These differences are much likely be explained by the fact that - due to the way the peak finding algorithm works - no strict correlation between the intensity threshold (algorithm) and the threshold which was manually set could be derived (see supplemental explanation S2). Based on these data, the intensity threshold of 100 cps which has manually been applied seems to be lower than the one used by the algorithm. This is partly attributed to the background subtraction used during the peak finding but also due to the lack of comparability between parameter settings. Again, the recovery rates in the wastewater matrix indicate the strongest matrix effects, for instance resulting in a recovery rate of about 54% (67 out of 123) at 25 ng L^{-1}. The second problem when processing real data is indicated by the false positive candidates, predominantly occurring in the most complex matrix and at the lowest concentration level. If disregarding the wastewater matrix, on average, the fast majority of 96, 115 and 121 out of 123 spiked standards could automatically be detected and 25, 100 and 500 ng L^{-1}. These findings demonstrate the general capability of the method.

When assessing the different models, it is conspicuous that the numbers of recognized standards do not substantially deviate from the ones obtained by the peak finding algorithm suggesting that the peak finding itself is the limiting factor of the whole procedure. Using an even lower non-target intensity threshold would increase the number of recognized standards. However, our experience has shown that lower intensity thresholds produce other problems (e.g. substantially higher number of false positives) which are therefore not recommended. An alternative strategy might be the combination of different peak finding algorithms to pool the individual strengths and therefore increase the feature coverage before any filter criteria were applied. This, however, was not within the scope of this study and has therefore not been investigated further.

One exception represents model (e) which requires three different injection volumes. As the minimum volume was only 50% of the normally injected one, it is not surprising to sacrifice sensitivity and hence decreasing the number of standards above the threshold. The fact that the effective concentration gets halved is the main reason why the numbers of recorded standards for model (e), compared to other models, always represent the lowest recovery rate. Notwithstanding model (e), the absolute number of substances (based on the non-target peak finding) filtered by the models did not exceed eight (see Table 4a, 25 ng L^{-1}, model (d), pretreated wastewater). The average number of features removed was less than two leading to a mean recovery rate of more than 98% (more than 96% if also considering model (e)) with respect to the number of peaks detected by the peak finding algorithm. The aforementioned relationships, i.e. better recovery rates the higher the concentration as well as problematics caused by matrix interferences can also be observed after applying the models. The mean recovery rate based on the total number of 123, over all models and matrix types, was higher than 67%, 88% and 96% for 25 ng L^{-1}, 100 ng L^{-1} and 500 ng L^{-1}, respectively. Considering the number obtained by the non-target peak finding as a reference, rates of more than 94%, 97% and 99% would be reached. It is noteworthy, that it is theoretically possible, that the number of

standards taken from the respective model is higher than the number of features recognized by the peak finding algorithm (see Table 4a, 500 ng L^{-1} river water). This phenomenon occurs in cases where a real peak was just partially recognized over four replicates (false negative), however, as a result of the subsequent peak integration, the feature can be corrected and, due to the combinatorics (OR linkage), remain as a real peak (not possible for model (a) as only one injection is regarded). Taking into account the maximum achievable number of hits for the respective model (model (x) manual), more than 90% waiving model (e) as well as the wastewater matrix were retained after applying the different models.

Negative ionization mode

The general statement for the negative ionization mode is very similar to the above mentioned one. The set point of $N_{s-} = 56$ was reached in ultrapure water at a concentration of 500 ng L^{-1}. In contrast to the positive ionization mode, the negative one was generally less sensitive leading to worse recovery rates if regarding same concentrations and resulting, on average, in a decrease in the recovery of about 28%. This circumstance was already recognized during the manual peak finding step, for instance in case of 100 ng L^{-1} 44, 43 and 40 out of 56 standards could manually be confirmed in ultrapure water (-E-), groundwater (-F-) and river water (-G-) while only 27 ones could be verified in the pretreated wastewater (-H-). Disregarding the wastewater matrix, 97% of the 56 standards could on average be confirmed at 500 ng L^{-1} while more than three-quarters were recorded at 100 ng L^{-1}. At the lowest concentration level, however, less than half of the spiked standards could be detected. Based on the numbers that could manually be confirmed, the peak finding algorithm reaches mean recovery rates of 54%, 67% and 90% for 25 ng L^{-1}, 100 ng L^{-1} and 500 ng L^{-1}, respectively which are significantly lower compared to the positive mode (82%, 96% and 98%). These results at least partially stem from the fact that the sensitivity, hence the intensity in general was lower. The substantially poorer sensitivity in the negative ionization mode is partly due to the fact that, in order to assure comparable retention times between the two ionization modes, formic acid was used as eluent additive in the negative mode as well. The implementation of different intensity thresholds for positive and negative ionization mode [2] would lead to lower discrepancies between both modes. Alternatively, a signal-to-noise based threshold instead of a fixed intensity threshold should be preferred.

These findings once more underscore the challenges of the actual peak finding, but do not represent a restriction with regard to the evaluation of the models since merely the maximum achievable number of features is affected. The lowest remaining number of standards recognized can be identified for model (e) for the same reasons as already discussed above. Disregarding model (e), the maximum number of standards removed by the filters was six, whereas the mean number of features removed was less than two (based on the non-target peak finding). Considering all models with respect to the numbers obtained by the non-target peak finding, a mean recovery rate of more than 91% is reached. The mean recovery rates of 19%, 42% and 78% for 25 ng L^{-1}, 100 ng L^{-1} and 500 ng L^{-1} appear to be apparently low, however, in relation to the non-target results, rates of 90%, 92% and 94% can be discerned. The mean recovery rate with respect to the maximum achievable number of the respective model (model (x) manual) was 89% whereby the minimum rate was 71%. These findings clearly demonstrate that the application of the models only lead to a very limited elimination of the spiked standards. Furthermore, the influence of the concentration and the sample matrix could be shown. As expected, most problems occur in complex samples at low concentration levels leading to a significant reduction of the

number of recognized standards. This, however, already appears during peak finding whereas the application of the presented models does not have a major impact on the standards. The overall recognitions, with respect to the recognized standards, still lie within a satisfactory range. Please note that all results were obtained without any correction of the peak integration (which was optimized in a former study, data not shown). When disregarding the wastewater matrix, about three-quarters of the spiked standards could be detected at a concentration level of 100 ng L^{-1}. This indicates that for several compounds, the health orientation value of 100 ng L^{-1} as suggested by the German Federal Environment Agency [18,19] is difficult to be reached in the negative ionization mode which is much likely attributed to the fact that formic acid has been used as eluent additive in the negative mode as well.

5. Conclusions

This study focuses on the comparison of different data evaluation strategies of LC-HRMS data. The assessment is based on a novel combinatorial approach using different numbers of technical replicates. The repeatability calculated for all features laying bare the weaknesses in peak finding and thus emphasizes the need of sophisticated processing strategies for dealing with false positive and false negative findings. We found that the use of replicates in combination with stringent filter criteria clearly led to an enhance data quality since both the repeatability and the peak recognition could be considerable improved. Reference standards which were spiked as a control were barely or not removed by these measures. For the first time, the problem of false negative findings was regarded on the basis of unknown features in addition to spiked reference standards.

On the basis of the presented data, it is strongly recommended to use at least two, or better, three technical replicates of each sample. We believe model (c) reflects an acceptable compromise between the benefits of reducing false positives and correcting partially false negatives and the drawback of the lower sample throughput. As most of the challenges are attributed to weaknesses in peak finding, the combination of different peak finding algorithms could be useful to increase the feature coverage.

Note that these general strategies are transferable to different data types regardless of the software tools which have been used for the feature extraction (vendors or open source software packages). It should, however, be recalled that many peak finding algorithms comprise the subsequent peak integration directly in the workflow which is why the problem of partially false negative findings is concealed and can only be identified if processing replicates individually.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.aca.2016.06.030.

References

[1] M. Krauss, H. Singer, J. Hollender, LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns, Anal. Bioanal. Chem. 397 (2010) 943–951.
[2] P. Gago-Ferrero, et al., Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewater with LC-HRMS/MS, Environ. Sci. Technol. 49 (20) (2015) 12333–12341.
[3] C. Hug, et al., Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening, Environ. Pollut. 184 (2014) 25–32.
[4] E.L. Schymanski, et al., Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry, Environ. Sci. Technol. 48 (3) (2013) 1811–1818.
[5] E. Schymanski, et al., Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis, Anal. Bioanal. Chem. 407 (21) (2015) 6237–6255.
[6] L. Vergeynst, H. Van Langenhove, K. Demeestere, Balancing the false negative and positive rates in suspect screening with high-resolution orbitrap mass spectrometry using multivariate statistics, Anal. Chem. 87 (4) (2015) 2170–2177.
[7] E. Tengstrand, et al., A concept study on non-targeted screening for chemical contaminants in food using liquid chromatography–mass spectrometry in combination with a metabolomics approach, Anal. Bioanal. Chem. 405 (4) (2013) 1237–1243.
[8] M. Katajamaa, M. Orešič, Data processing for mass spectrometry-based metabolomics, J. Chromatogr. A 1158 (1–2) (2007) 318–328.
[9] J. Zhang, et al., Review of peak detection algorithms in liquid-chromatography-mass spectrometry, Curr. Genomics 10 (6) (2009) 388–401.
[10] J.B. Coble, C.G. Fraga, Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery, J. Chromatogr. A 1358 (2014) 155–164.
[11] L. Vergeynst, et al., Suspect screening and target quantification of multi-class pharmaceuticals in surface water based on large-volume injection liquid chromatography and time-of-flight mass spectrometry, Anal. Bioanal. Chem. 406 (11) (2014) 2533–2547.
[12] W. Yu, et al., Improving mass spectrometry peak detection using multiple peak alignment results, J. Proteome Res. 7 (1) (2008) 123–129.
[13] T. Yu, D.P. Jones, Improving peak detection in high-resolution LC/MS metabolomics data using preexisting knowledge and machine learning approach, Bioinformatics 30 (20) (2014) 2941–2948.
[14] C.A. Smith, et al., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (3) (2006) 779–787.
[15] L. Brodsky, et al., Evaluation of peak picking quality in LC−MS metabolomics data, Anal. Chem. 82 (22) (2010) 9177–9187.
[16] A. Müller, et al., A new approach to data evaluation in the non-target screening of organic trace substances in water analysis, Chemosphere 85 (8) (2011) 1211–1219.
[17] C. Ort, et al., Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis, Addiction 109 (8) (2014) 1338–1352.
[18] A. Bergmann, et al., Potential water-related environmental risks of hydraulic fracturing employed in exploration and exploitation of unconventional natural gas reservoirs in Germany, Environ. Sci. Eur. 26 (1) (2014) 1–14.
[19] Umweltbundesamt, Bewertung der anwesenheit teil- oder nicht bewertbarer stoffe im trinkwasser aus gesundheitlicher sicht, Bundesgesundheitsblatt - Gesundheitsforsch. - Gesundheitsschutz 46 (3) (2003) 249–251.
[20] Sciex, MarkerView™ Software - Reference Manual, 2010.
[21] J. von Neumann, Distribution of the Ratio of the Mean Square Successive Difference to theVariance, 4, 1941, pp. 367–395.

Supplementary content to:

General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry

Tobias Bader[1,2*], Wolfgang Schulz[1], Klaus Kümmerer[2] and Rudi Winzenbacher[1]

[1]*Laboratory for Operation Control and Research, Zweckverband Landeswasserversorgung, Am Spitzigen Berg 1, 89129 Langenau, Germany*

[2]*Sustainable Chemistry and Material Resources, Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg, Scharnhorststraße 1/C13, 21335 Lüneburg, Germany*

*Corresponding author
Tel.: +49 7345 9638 2279, E-mail address: Bader.T@lw-online.de

****

Table of contents

3.2 System stability

3.3 Processing models

3.4 Characteristics derived for the comprehensive evaluation

4. Results and discussion

4.1 Comprehensive evaluation

4.2. Final confirmation

5. Conclusions

6. Acknowledgements

7. References

****

Table S1

Stock solution I: Pharmaceutical drugs (N = 32)

#	Name	Elemental composition	Retention time [min]
1	10,11-Dihydro-10,11-dihydroxy carbamazepine	$C_{15}H_{14}N_2O_3$	6.0
2	2-Ethyl-2-phenylmalonamide	$C_{11}H_{14}N_2O_2$	5.3
3	Atenolol	$C_{14}H_{22}N_2O_3$	4.5
4	Betaxolol	$C_{18}H_{29}NO_3$	7.0
5	Bezafibrate	$C_{19}H_{20}NO_4Cl$	10.4
6	Bisoprolol	$C_{18}H_{31}NO_4$	6.4
7	Carbamazepine	$C_{15}H_{12}N_2O$	8.4
8	Carbamazepine-10,11 epoxide	$C_{15}H_{12}N_2O_2$	7.2
9	Dapsone	$C_{12}H_{12}N_2O_2S$	6.3
10	Diazepam	$C_{16}H_{13}N_2OCl$	10.7
11	Diclofenac	$C_{14}H_{11}NO_2Cl_2$	12.0
12	Etofibrate	$C_{18}H_{18}NO_5Cl$	12.3
13	Fenofibrate	$C_{20}H_{21}O_4Cl$	15.9
14	Fenofibric acid	$C_{17}H_{15}O_4Cl$	12.0
15	Iopamidol	$C_{17}H_{22}I_3N_3O_8$	4.4
16	Ketoprofen	$C_{16}H_{14}O_3$	10.1
17	Metoprolol	$C_{15}H_{25}NO_3$	5.5
18	Metronidazole	$C_6H_9N_3O_3$	4.7
19	N-Acetyl sulfamethoxazole	$C_{12}H_{13}N_3O_4S$	6.9
20	Pentoxifylline	$C_{13}H_{18}N_4O_3$	6.1
21	Phenacetin	$C_{10}H_{13}NO_2$	7.2
22	Phenazone	$C_{11}H_{12}N_2O$	5.8
23	Pindolol	$C_{14}H_{20}N_2O_2$	5.0
24	Primidone	$C_{12}H_{14}N_2O_2$	5.8
25	Propranolol	$C_{16}H_{21}NO_2$	6.7
26	Ronidazole	$C_6H_8N_4O_4$	5.0
27	Sotalol	$C_{12}H_{20}N_2O_3S$	4.5
28	Sulfadiazine	$C_{10}H_{10}N_4O_2S$	5.0
29	Sulfadimidine	$C_{12}H_{14}N_4O_2S$	5.8
30	Sulfamerazine	$C_{11}H_{12}N_4O_2S$	5.5
31	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	6.8
32	Trimethoprim	$C_{14}H_{18}N_4O_3$	5.0

Table S2

Stock solution II: Multi-component standard (N = 263)

#	Name	Elemental composition	Retention time [min]
1	(Benzothiazol-2-ylthio)methylthiocyanat	$C_9H_6N_2S_3$	11.9
2	2-(Methylthio)benzothiazole	$C_8H_7NS_2$	11.5
3	2,6-Dichlorbenzamid	$C_7H_5Cl_2NO$	5.8
4	2-Aminobenzothiazole	$C_7H_6N_2S$	4.8
5	2-Hydroxybenzothiazole	C_7H_5NOS	7.3
6	4′-Hydroxydiclofenac	$C_{14}H_{11}Cl_2NO_3$	9.8
7	5,6-Dimethyl-1H-benzotriazole	$C_8H_9N_3$	7.3
8	Acetyl-sulfamethoxazole	$C_{12}H_{13}N_3O_4S$	6.9
9	Aclonifen	$C_{12}H_9ClN_2O_3$	13.2
10	Alachlor	$C_{14}H_{20}ClNO_2$	12.7
11	Amantadine	$C_{10}H_{17}N$	5.0
12	Ametryn	$C_9H_{17}N_5S$	8.1
13	Amidosulfuron	$C_9H_{15}N_5O_7S_2$	9.3
14	Amisulpride	$C_{17}H_{27}N_3O_4S$	5.1
15	Amisulpride N-Oxide	$C_{17}H_{27}N_3O_5S$	5.2
16	Aspartame	$C_{14}H_{18}N_2O_5$	5.2
17	Atenolol	$C_{14}H_{22}N_2O_3$	4.5
18	Atraton	$C_9H_{17}N_5O$	6.2
19	Atrazine	$C_8H_{14}ClN_5$	9.6
20	Atrazine-2-hydroxy	$C_8H_{15}N_5O$	5.0
21	Azinphos-methyl	$C_{10}H_{12}N_3O_3PS_2$	11.2
22	Azoxystrobin	$C_{22}H_{17}N_3O_5$	11.7
23	Benalaxyl-M	$C_{20}H_{23}NO_3$	13.4
24	Benazolin	$C_9H_6ClNO_3S$	8.0
25	Bensulfuron-methyl	$C_{16}H_{18}N_4O_7S$	10.4
26	Bentazon	$C_{10}H_{12}N_2O_3S$	9.3
27	Benzothiazole-6-carboxylic acid	$C_8H_5NO_2S$	6.1
28	Benzotriazole	$C_6H_5N_3$	5.5
29	Betaxolol	$C_{18}H_{29}NO_3$	7.0
30	Bezafibrate	$C_{19}H_{20}ClNO_4$	10.3
31	Bisoprolol	$C_{18}H_{31}NO_4$	6.4
32	Boscalid	$C_{18}H_{12}Cl_2N_2O$	11.9
33	Bromacil	$C_9H_{13}BrN_2O_2$	8.0
34	Caffeine	$C_8H_{10}N_4O_2$	5.0
35	Candesartan	$C_{24}H_{20}N_6O_3$	9.2
36	Carbamazepine	$C_{15}H_{12}N_2O$	8.4
37	Carbamazepine 10,11-epoxide	$C_{15}H_{12}N_2O_2$	7.2
38	Carbendazim	$C_9H_9N_3O_2$	4.8
39	Carbetamide	$C_{12}H_{16}N_2O_3$	7.8
40	Carbofuran	$C_{12}H_{15}NO_3$	9.1
41	Chloramben	$C_7H_5Cl_2NO_2$	7.1
42	Chloramphenicol	$C_{11}H_{12}Cl_2N_2O_5$	7.1
43	Chlorfenvinphos	$C_{12}H_{14}Cl_3O_4P$	13.2
44	Chloridazon	$C_{10}H_8ClN_3O$	6.4
45	Chloridazon-methyl-desphenyl	$C_5H_6ClN_3O$	4.6
46	Chlorotoluron	$C_{10}H_{13}ClN_2O$	9.3
47	Chloroxuron	$C_{15}H_{15}Cl_2N_2O$	11.4
48	Chlorpyrifos	$C_9H_{11}Cl_3NO_3PS$	15.8
49	Chlortetracycline	$C_{22}H_{23}ClN_2O_8$	5.5
50	Chlorthalonil R611965	$C_8H_4Cl_3NO_3$	5.4
51	Clarithromycin	$C_{38}H_{69}NO_{13}$	8.4
52	Clenbuterol	$C_{12}H_{18}Cl_2N_2O$	5.6
53	Clomazone	$C_{12}H_{14}ClNO_2$	10.7
54	Clopyralid	$C_6H_3Cl_2NO_2$	5.3
55	Clothianidin	$C_6H_8ClN_5O_2S$	6.2
56	Codeine	$C_{18}H_{21}NO_3$	4.7
57	Crotamiton	$C_{13}H_{17}NO$	10.6
58	Cyanazine	$C_9H_{13}ClN_6$	8.4
59	Cyproconazole	$C_{15}H_{18}ClN_3O$	11.1
60	Cyprodinil	$C_{14}H_{15}N_3$	10.6
61	Dapsone	$C_{12}H_{12}N_2O_2S$	6.3
62	DEET	$C_{12}H_{17}NO$	9.7
63	Deisopropylatrazine	$C_5H_8ClN_5$	5.5
64	Desethylatrazine	$C_6H_{10}ClN_5$	6.5
65	Desethylterbutylazin	$C_7H_{12}ClN_5$	8.5
66	Desmetryn	$C_8H_{15}N_5S$	7.0
67	Diatrizoic acid	$C_{11}H_9I_3N_2O_4$	4.3
68	Diazepam	$C_{16}H_{13}ClN_2O$	10.6

4

#	Name	Elemental composition	Retention time [min]
69	Diazinon	$C_{12}H_{21}N_2O_3PS$	14.0
70	Dichlorvos	$C_4H_7Cl_2O_4P$	8.4
71	Diclofenac	$C_{14}H_{11}Cl_2NO_2$	12.0
72	Difenoconazole	$C_{19}H_{17}Cl_2N_3O_3$	13.4
73	Diflubenzuron	$C_{14}H_9ClF_2N_2O_2$	12.3
74	Diflufenican	$C_{19}H_{11}F_5N_2O_2$	14.5
75	Dihydrocodeine	$C_{18}H_{23}NO_3$	4.6
76	Dimefuron	$C_{15}H_{19}ClN_4O_3$	10.6
77	Dimethachlor	$C_{13}H_{18}ClNO_2$	10.5
78	Dimethachlor CGA 102935	$C_{12}H_{13}NO_5$	5.5
79	Dimethachlor CGA 354742	$C_{13}H_{19}NO_5S$	6.0
80	Dimethachlor CGA 369873	$C_{10}H_{13}NO_4S$	5.1
81	Dimethachlor CGA 373464	$C_{12}H_{15}NO_6S$	5.3
82	Dimethachlor CGA 50266	$C_{13}H_{17}NO_4$	6.8
83	Dimethachlor SYN 528702	$C_{15}H_{21}NO_5S$	6.8
84	Dimethachlor SYN 530561	$C_{13}H_{17}NO_5$	6.0
85	Dimethenamid	$C_{12}H_{18}ClNO_2S$	11.5
86	Dimethenamid M23	$C_{12}H_{17}NO_4S$	7.5
87	Dimethenamid M27	$C_{12}H_{19}NO_5S_2$	6.5
88	Dimethenamid-P M31	$C_{14}H_{21}NO_5S_2$	7.4
89	Dimethoate	$C_5H_{12}NO_3PS_2$	6.7
90	Dimoxystrobin	$C_{19}H_{22}N_2O_3$	12.8
91	Dimoxystrobin 505M08 (BF 505-7)	$C_{19}H_{20}N_2O_5$	9.3
92	Dimoxystrobin 505M09 (BF 505-8)	$C_{19}H_{20}N_2O_5$	9.7
93	Diuron	$C_9H_{10}Cl_2N_2O$	9.8
94	Doxycycline	$C_{22}H_{24}N_2O_8$	6.2
95	Epoxiconazole	$C_{17}H_{13}ClFN_3O$	11.8
96	Eprosartan	$C_{23}H_{24}N_2O_4S$	6.5
97	Erythromycin	$C_{37}H_{67}NO_{13}$	7.3
98	Ethidimuron	$C_7H_{12}N_4O_3S_2$	6.3
99	Ethofumesate	$C_{13}H_{18}O_5S$	12.2
100	Etofibrate	$C_{18}H_{18}ClNO_5$	12.2
101	Fenhexamid	$C_{14}H_{17}Cl_2NO_2$	11.9
102	Fenofibrate	$C_{20}H_{21}ClO_4$	15.8
103	Fenoxaprop	$C_{16}H_{12}ClNO_5$	11.9
104	Fenoxycarb	$C_{17}H_{19}NO_4$	12.6
105	Fenpropidin	$C_{19}H_{31}N$	8.9
106	Fenpropimorph	$C_{20}H_{33}NO$	9.0
107	Fenuron	$C_9H_{12}N_2O$	6.4
108	Flamprop	$C_{16}H_{13}ClFNO_3$	10.4
109	Flazasulfuron	$C_{13}H_{12}F_3N_5O_5S$	10.4
110	Florasulam	$C_{12}H_8F_3N_5O_3S$	9.0
111	Fluazifop	$C_{15}H_{12}F_3NO_4$	11.1
112	Flufenacet	$C_{14}H_{13}F_4N_3O_2S$	12.8
113	Flufenacet-FOE 5043	$C_{11}H_{14}FNO_4S$	6.1
114	Fluopicolide	$C_{14}H_8Cl_3F_3N_2O$	12.2
115	Fluroxypyr	$C_7H_5Cl_2FN_2O_3$	8.1
116	Flurtamone	$C_{18}H_{14}F_3NO_2$	11.2
117	Flusilazole	$C_{16}H_{15}F_2N_3Si$	12.2
118	Foramsulfuron	$C_{17}H_{20}N_6O_7S$	8.0
119	Gabapentin	$C_9H_{17}NO_2$	4.7
120	Gabapentin-lactam	$C_9H_{15}NO$	7.2
121	Haloxyfop	$C_{15}H_{11}ClF_3NO_4$	12.4
122	Hexazinone	$C_{12}H_{20}N_4O_2$	7.7
123	Imidacloprid	$C_9H_{10}ClN_5O_2$	6.5
124	Indometacin	$C_{19}H_{16}ClNO_4$	12.0
125	Iodosulfuron-methyl	$C_{14}H_{14}IN_5O_6S$	10.6
126	Iohexol	$C_{19}H_{26}I_3N_3O_9$	4.3
127	Iomeprol	$C_{17}H_{22}I_3N_3O_8$	4.4
128	Iopamidol	$C_{17}H_{22}I_3N_3O_8$	3.7
129	Irbesartan	$C_{25}H_{28}N_6O$	8.8
130	Irbesartan_446	$C_{25}H_{30}N_6O_2$	9.7
131	Irgarol	$C_{11}H_{19}N_5S$	9.8
132	Irgarol-descyclopropyl	$C_8H_{15}N_5S$	7.1
133	iso-Chloridazon	$C_{10}H_8ClN_3O$	8.3
134	Isoproturon	$C_{12}H_{18}N_2O$	9.7
135	Ketoprofen	$C_{16}H_{14}O_3$	10.1
136	Kresoxim-methyl	$C_{18}H_{19}NO_4$	13.4
137	Lamotrigine	$C_9H_7Cl_2N_5$	5.4
138	Lamotrigine N2-Oxide	$C_9H_7Cl_2N_5O$	5.6
139	Linuron	$C_9H_{10}Cl_2N_2O_2$	11.3
140	Losartan	$C_{22}H_{23}ClN_6O$	9.1

#	Name	Elemental composition	Retention time [min]
141	Malathion	$C_{10}H_{19}O_6PS_2$	12.5
142	Mesosulfuron-methyl	$C_{17}H_{21}N_5O_9S_2$	9.5
143	Metalaxyl	$C_{15}H_{21}NO_4$	9.8
144	Metalaxyl-M CGA 108906	$C_{14}H_{17}NO_6$	6.4
145	Metalaxyl-M CGA 62826/NOA 409045 (R-form)	$C_{14}H_{19}NO_4$	8.1
146	Metamitron	$C_{10}H_{10}N_4O$	6.1
147	Metazachlor	$C_{14}H_{16}ClN_3O$	10.4
148	Metazachlor BH 479-11	$C_{15}H_{19}N_3O_2S$	6.9
149	Metazachlor BH 479-12	$C_{14}H_{13}N_3O_5$	5.3
150	Metazachlor BH 479-4	$C_{14}H_{15}N_3O_3$	6.2
151	Metazachlor BH 479-8	$C_{14}H_{17}N_3O_4S$	5.9
152	Metazachlor BH 479-9	$C_{16}H_{19}N_3O_4S$	6.5
153	Methabenzthiazuron	$C_{10}H_{11}N_3OS$	9.1
154	Methidathion	$C_6H_{11}N_2O_4PS_3$	11.2
155	Methylphenidat	$C_{14}H_{19}NO_2$	5.6
156	Metobromuron	$C_9H_{11}BrN_2O_2$	10.1
157	Metolachlor	$C_{15}H_{22}ClNO_2$	12.7
158	Metolachlor CGA 354743 / CGA 380168 (S-form)	$C_{15}H_{23}NO_5S$	7.2
159	Metolachlor CGA 51202 / CGA 351916 (S-form)	$C_{15}H_{21}NO_4$	9.2
160	Metoprolol	$C_{15}H_{25}NO_3$	5.5
161	Metoprolol acid	$C_{14}H_{21}NO_4$	4.8
162	Metosulam	$C_{14}H_{13}Cl_2N_5O_4S$	9.6
163	Metoxuron	$C_{10}H_{13}ClN_2O_2$	7.7
164	Metribuzin	$C_8H_{14}N_4OS$	8.6
165	Metronidazole	$C_6H_9N_3O_3$	4.7
166	Metsulfuron-methyl	$C_{14}H_{15}N_5O_6S$	8.8
167	Mevinphos	$C_7H_{13}O_6P$	6.3
168	Monolinuron	$C_9H_{11}ClN_2O_2$	9.8
169	Monuron	$C_9H_{11}ClN_2O$	8.2
170	N-Acetyl-4-aminoantipyrine	$C_{13}H_{15}N_3O_2$	5.1
171	Nadolol	$C_{17}H_{27}NO_4$	4.8
172	Napropamide	$C_{17}H_{21}NO_2$	12.2
173	Naproxen	$C_{14}H_{14}O_3$	10.2
174	N-Formyl-4-aminoantipyrine	$C_{12}H_{13}N_3O_2$	5.1
175	Nicosulfuron	$C_{15}H_{18}N_6O_6S$	7.7
176	Olmesartan	$C_{24}H_{26}N_6O_3$	6.6
177	Oxadixyl	$C_{14}H_{18}N_2O_4$	8.1
178	Oxazepam	$C_{15}H_{11}ClN_2O_2$	8.8
179	Oxytetracycline	$C_{22}H_{24}N_2O_9$	5.1
180	Parathion	$C_{10}H_{14}NO_5PS$	13.6
181	Parathion-methyl	$C_8H_{10}NO_5PS$	11.9
182	Pendimethalin	$C_{13}H_{19}N_3O_4$	15.8
183	Pentoxifylline	$C_{13}H_{18}N_4O_3$	6.1
184	Pethoxamid	$C_{16}H_{22}ClNO_2$	12.6
185	Phenacetin	$C_{10}H_{13}NO_2$	7.2
186	Phenazone	$C_{11}H_{12}N_2O$	5.8
187	Phenylethylmalonamide	$C_{11}H_{14}N_2O_2$	5.3
188	Phosalone	$C_{12}H_{15}ClNO_4PS_2$	14.3
189	Phoxim	$C_{12}H_{15}N_2O_3PS$	14.3
190	Picloram	$C_6H_3Cl_3N_2O_2$	5.9
191	Picolinafen	$C_{19}H_{12}F_4N_2O_2$	15.1
192	Picoxystrobin	$C_{18}H_{16}F_3NO_4$	13.5
193	Pindolol	$C_{14}H_{20}N_2O_2$	5.0
194	Pirimicarb	$C_{11}H_{18}N_4O_2$	5.9
195	Primidone	$C_{12}H_{14}N_2O_2$	5.8
196	Primisulfuron-methyl	$C_{15}H_{12}F_4N_4O_7S$	11.9
197	Prochloraz	$C_{15}H_{16}Cl_3N_3O_2$	11.0
198	Prometon	$C_{10}H_{19}N_5O$	7.0
199	Propazine	$C_9H_{16}ClN_5$	10.9
200	Propazine-2-hydroxy	$C_9H_{17}N_5O$	5.4
201	Propiconazole	$C_{15}H_{17}Cl_2N_3O_2$	12.8
202	Propyphenazone	$C_{14}H_{18}N_2O$	9.1
203	Prosulfocarb	$C_{14}H_{21}NOS$	15.0
204	Prosulfuron	$C_{15}H_{16}F_3N_5O_4S$	11.2
205	Quinmerac	$C_{11}H_8ClNO_2$	6.6
206	Quinmerac BH 518-2	$C_{11}H_6ClNO_4$	5.8
207	Quinoxyfen	$C_{15}H_8Cl_2FNO$	14.8
208	Rimsulfuron	$C_{14}H_{17}N_5O_7S_2$	9.1
209	Ritalinic acid	$C_{13}H_{17}NO_2$	5.1
210	Ronidazole	$C_6H_8N_4O_4$	4.9
211	Roxithromycin	$C_{41}H_{76}N_2O_{15}$	8.5
212	Sebuthylazine	$C_9H_{16}ClN_5$	10.7

#	Name	Elemental composition	Retention time [min]
213	Sebuthylazine-desethyl	$C_7H_{12}ClN_5$	7.6
214	Simazine	$C_7H_{12}ClN_5$	8.3
215	S-Metolachlor CGA 357704	$C_{14}H_{17}NO_5$	7.1
216	S-Metolachlor CGA 368208	$C_{11}H_{15}NO_4S$	5.7
217	S-Metolachlor CGA 50267	$C_{12}H_{17}NO_2$	8.6
218	S-Metolachlor CGA 50720	$C_{11}H_{13}NO_3$	6.5
219	S-Metolachlor CGA 37735	$C_{11}H_{15}NO_2$	6.8
220	S-Metolachlor NOA 413173	$C_{14}H_{19}NO_6S$	6.5
221	Sotalol	$C_{12}H_{20}N_2O_3S$	4.5
222	Spiroxamine	$C_{18}H_{35}NO_2$	9.1
223	Sulfadiazine	$C_{10}H_{10}N_4O_2S$	5.0
224	Sulfadimethoxine	$C_{12}H_{14}N_4O_4S$	7.6
225	Sulfadimidine	$C_{12}H_{14}N_4O_2S$	5.8
226	Sulfadoxine	$C_{12}H_{14}N_4O_4S$	6.7
227	Sulfamerazine	$C_{11}H_{12}N_4O_2S$	5.4
228	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	6.7
229	Sulfathiazole	$C_9H_9N_3O_2S_2$	5.1
230	Sulpirid	$C_{15}H_{23}N_3O_4S$	4.6
231	Sulpiride N-Oxide	$C_{15}H_{23}N_3O_5S$	4.6
232	Tebuconazole	$C_{16}H_{22}ClN_3O$	12.1
233	Tebutam	$C_{15}H_{23}NO$	12.7
234	Telmisartan	$C_{33}H_{30}N_4O_2$	8.3
235	Terbuthylazine	$C_9H_{16}ClN_5$	11.2
236	Terbuthylazine 1 SYN 545666	$C_8H_{14}N_4O_2$	5.3
237	Terbuthylazine-desethyl-2-hydroxy	$C_7H_{13}N_5O$	4.6
238	Terbutryn	$C_{10}H_{19}N_5S$	9.3
239	Terbutylazine 2 CGA 324007	$C_7H_{12}N_4O_2$	4.9
240	Tetracycline	$C_{22}H_{24}N_2O_8$	5.2
241	Thiacloprid	$C_{10}H_9ClN_4S$	7.5
242	Thiacloprid sulfonic acid, M30; BCS-AB54351	$C_{10}H_{13}ClN_4O_5S$	4.9
243	Thiamethoxam	$C_8H_{10}ClN_5O_3S$	5.7
244	Thifensulfuron-methyl (O-desmethyl-thifensulfuronmethyl, IN-L9226)	$C_{12}H_{13}N_5O_6S_2$	8.5
245	Tolyltriazole	$C_7H_7N_3$	6.4
246	Topramezone	$C_{16}H_{17}N_3O_5S$	6.1
247	Tramadol	$C_{16}H_{25}NO_2$	5.6
248	Tramadol N-Oxide	$C_{16}H_{25}NO_3$	5.7
249	trans-10,11-Dihydro-10,11-dihydroxy Carbamazepine	$C_{15}H_{14}N_2O_3$	6.0
250	Triadimenol	$C_{14}H_{18}ClN_3O_2$	10.8
251	Triallate	$C_{10}H_{16}Cl_3NOS$	16.4
252	Triasulfuron	$C_{14}H_{16}ClN_5O_5S$	9.1
253	Triclopyr	$C_7H_4Cl_3NO_3$	10.2
254	Trifloxystrobin	$C_{20}H_{19}F_3N_2O_4$	14.7
255	Trifloxystrobin CGA 321113	$C_{19}H_{17}F_3N_2O_4$	12.8
256	Trifloxystrobin NOA 413161	$C_{19}H_{15}F_3N_2O_6$	10.1
257	Triflusulfuron-methyl	$C_{17}H_{19}F_3N_6O_6S$	11.9
258	Trimethoprim	$C_{14}H_{18}N_4O_3$	5.0
259	Tritosulfuron	$C_{13}H_9F_6N_5O_4S$	11.6
260	Tritosulfuron BH 635-4 (635M01)	$C_{10}H_{10}F_3N_5O_4S$	5.8
261	Tritosulfuron BH 635-5	$C_5H_5F_3N_4O$	6.9
262	Valsartan	$C_{24}H_{29}N_5O_3$	10.7
263	Valsartan acid	$C_{14}H_{10}N_4O_2$	7.3

Table S3

Suppliers of chemical standards

#	Company	Location
1	Alfa Aesar	Karlsruhe, Germany
2	BASF	Ludwigshafen, Germany
3	Bayer	Leverkusen, Germany
4	BIOZOL Diagnostica	Eching, Germany
5	Campro Scientific	Berlin, Germany
6	CHEMOS	Regenstauf, Germany
7	ChiroBlock	Bitterfeld-Wolfen, Germany
8	Dr. Ehrenstorfer	Augsburg, Germany
9	European Directorate for the Quality of Medicines	Strasbourg, France
10	LGC Standards	Wesel, Germany
11	Molekula	Munich, Germany
12	Neochema	Bodenheim, Germany
13	Sigma-Aldrich	Steinheim am Albuch, Germany
14	Syngenta	Basel, Switzerland
15	Toronto Research Chemicals	Toronto, Canada
16	United States Pharmacopeia	Rockville, USA

Table S4

Stock solution III: Isotopically labeled standards (N = 130),
analyzed in both ESI modes (N$_{ESI+}$ = 123, N$_{ESI-}$ = 56)

#	Name	Elemental composition	Retention time [min]
1	2,4-D D3 [(-)]	C$_8$H$_3$^{2}H$_3$Cl$_2$O$_3$	9.7
2	2,6-Dichlorobenzamide-3,4,5 D3 [(+)]	C$_7$H$_2$^{2}H$_3$Cl$_2$NO	5.8
3	5-Methylbenzotriazole D6	C$_7$H^{2}H$_6$N$_3$	6.5
4	Acetyl-sulfamethoxazole D5	C$_{12}$H$_8$^{2}H$_5$N$_3$O$_4$S	7.0
5	Alachlor D13 [(+)]	C$_{14}$H$_7$^{2}H$_{13}$ClNO$_2$	12.8
6	Amisulpride D5	C$_{17}$H$_{22}$^{2}H$_5$N$_3$O$_4$S	5.1
7	Atazanavir D5	C$_{38}$H$_{47}$^{2}H$_5$N$_6$O$_7$	10.2
8	Atenolol acid D5	C$_{14}$H$_{16}$^{2}H$_5$NO$_4$	4.8
9	Atenolol D7 [(+)]	C$_{14}$H$_{15}$^{2}H$_7$N$_2$O$_3$	4.5
10	Atomoxetine D3 [(+)]	C$_{17}$H$_{18}$^{2}H$_3$NO	7.7
11	Atorvastatin D5	C$_{33}$H$_{30}$^{2}H$_5$FN$_2$O$_5$	11.8
12	Atrazine D5 [(+)]	C$_8$H$_9$^{2}H$_5$ClN$_5$	9.7
13	Atrazine-2-hydroxy D5	C$_8$H$_{10}$^{2}H$_5$N$_5$O	4.9
14	Atrazine-desisopropyl D5 [(+)]	C$_5$H$_3$^{2}H$_5$ClN$_5$	5.5
15	Azithromycin D3 [(+)]	C$_{38}$H$_{69}$^{2}H$_3$N$_2$O$_{12}$	5.8
16	Azoxystrobin D4 [(+)]	C$_{22}$H$_{13}$^{2}H$_4$N$_3$O$_5$	11.8
17	Bentazon D6	C$_{10}$H$_6$^{2}H$_6$N$_2$O$_3$S	9.4
18	Benzotriazole D4	C$_6$H^{2}H$_4$N$_3$	5.5
19	Bezafibrate D4	C$_{19}$H$_{16}$^{2}H$_4$ClNO$_4$	10.4
20	Bicalutamide D4	C$_{18}$H$_{10}$^{2}H$_4$F$_4$N$_2$O$_4$S	11.0
21	Caffeine D9 [(+)]	C$_8$H^{2}H$_9$N$_4$O$_2$	5.0
22	Candesartan D5	C$_{24}$H$_{15}$^{2}H$_5$N$_6$O$_3$	9.3
23	Carbamazepine D8 [(+)]	C$_{15}$H$_4$^{2}H$_8$N$_2$O	8.4
24	Carbamazepine-10,11-epoxide C13,D2 [(+)]	C$_{14}$^{13}CH$_{10}$^{2}H$_2$N$_2$O$_2$	7.2
25	Carbendazim D4 [(+)]	C$_9$H$_5$^{2}H$_4$N$_3$O$_2$	4.8
26	Cetirizine D8	C$_{21}$H$_{17}$^{2}H$_8$ClN$_2$O$_3$	8.3
27	Chloridazon D5	C$_{10}$H$_3$^{2}H$_5$ClN$_3$O	6.4
28	Chloridazon-methyl-desphenyl D3	C$_5$H$_3$^{2}H$_3$ClN$_3$O	4.5
29	Chlorotoluron D6 [(+)]	C$_{10}$H$_7$^{2}H$_6$ClN$_2$O	9.3
30	Chlorpyrifos D10 [(+)]	C$_9$H^{2}H$_{10}$Cl$_3$NO$_3$PS	15.9
31	Chlorpyrifos-methyl D6 [(+)]	C$_7$H^{2}H$_6$Cl$_3$NO$_3$PS	14.4
32	Citalopram D6 [(+)]	C$_{20}$H$_{15}$^{2}H$_6$FN$_2$O	7.3
33	Clarithromycin-N-methyl D3 [(+)]	C$_{38}$H$_{66}$^{2}H$_3$NO$_{13}$	8.4
34	Climbazole D4	C$_{15}$H$_{13}$^{2}H$_4$ClN$_2$O$_2$	8.4
35	Clofibric acid D4 [(-)]	C$_{10}$H$_7$^{2}H$_4$ClO$_3$	10.2
36	Clopidogrel carboxylic acid D4 [(+)]	C$_{15}$H$_{10}$^{2}H$_4$ClNO$_2$S	6.1
37	Clothianidin D3	C$_6$H$_5$^{2}H$_3$ClN$_5$O$_2$S	6.3
38	Clotrimazole D5 [(+)]	C$_{22}$H$_{12}$^{2}H$_5$ClN$_2$	8.7
39	Clozapine D8 [(+)]	C$_{18}$H$_{11}$^{2}H$_8$ClN$_4$	6.5
40	Codeine 13C,D3 [(+)]	C$_{17}$^{13}CH$_{18}$^{2}H$_3$NO$_3$	4.7
41	Cyclophosphamide D4 [(+)]	C$_7$H$_{11}$^{2}H$_4$Cl$_2$N$_2$O$_2$P	7.0
42	Cyprodinil D5 [(+)]	C$_{14}$^{2}H$_5$H$_{10}$N$_3$	10.7
43	Darunavir D9	C$_{27}$H$_{28}$^{2}H$_9$N$_3$O$_7$S	10.4
44	Desethylatrazine 15N3 [(+)]	C$_6$H$_{10}$ClN$_2$^{15}N$_3$	6.5
45	Desphenyl Chloridazon 15N2 [(+)]	C$_4$H$_4$ClN^{15}N$_2$O	2.9
46	Diazepam D5 [(+)]	C$_{16}$H$_8$^{2}H$_5$N$_2$OCl	10.7
47	Diazinon D10 [(+)]	C$_{12}$H$_{11}$^{2}H$_{10}$N$_2$O$_3$PS	14.1
48	Dichlorprop D6 [(-)]	C$_9$H$_2$^{2}H$_6$Cl$_2$O$_3$	10.7
49	Diclofenac D4	C$_{14}$H$_7$^{2}H$_4$Cl$_2$NO$_2$	12.1
50	Diflufenican D3	C$_{19}$H$_8$^{2}H$_3$F$_5$N$_2$O$_2$	14.7
51	Dimethenamid D3 [(+)]	C$_{12}$H$_{15}$^{2}H$_3$ClNO$_2$S	11.7
52	Dimethoate D6 [(+)]	C$_5$H$_6$^{2}H$_6$NO$_3$PS$_2$	6.7
53	Diuron D6	C$_9$H$_4$^{2}H$_6$Cl$_2$N$_2$O	9.8
54	Emtricitabine 13C,15N2 [(+)]	C$_7$^{13}CH$_{10}$FN^{15}N$_2$O$_3$S	4.5
55	Epoxiconazole D4 [(+)]	C$_{17}$H$_9$^{2}H$_4$ClFN$_3$O	11.9
56	Eprosartan D3	C$_{23}$H$_{21}$^{2}H$_3$N$_2$O$_4$S	6.6
57	Erythromycin 13C2 [(+)]	C$_{35}$^{13}C$_2$H$_{67}$NO$_{13}$	7.4
58	Fenofibrate D6 [(+)]	C$_{20}$H$_{15}$^{2}H$_6$ClO$_4$	15.9
59	Fipronil 13C2,15N2	C$_{10}$^{13}C$_2$H$_4$Cl$_2$F$_6$N$_2$^{15}N$_2$OS	13.4
60	Fluconazole D4	C$_{13}$H$_8$^{2}H$_4$F$_2$N$_6$O	5.9
61	Fluoxetine D5 [(+)]	C$_{17}$H$_{13}$^{2}H$_5$F$_3$NO	8.4
62	Furosemid D5 [(-)]	C$_{12}$H$_6$^{2}H$_5$ClN$_2$O$_5$S	8.3
63	Gabapentin D4	C$_9$H$_{13}$^{2}H$_4$NO$_2$	4.7
64	Hydrochlorothiazide 13C,D2	C$_6$^{13}CH$_6$^{2}H$_2$ClN$_3$O$_4$S$_2$	5.1
65	Ibuprofen D3 [(+)]	C$_{13}$H$_{15}$^{2}H$_3$O$_2$	12.4
66	Imidacloprid D4	C$_9$H$_6$^{2}H$_4$ClN$_5$O$_2$	6.5

9

#	Name	Elemental composition	Retention time [min]
67	Indomethacin D4	$C_{19}H_{12}{}^2H_4ClNO_4$	12.1
68	Irbesartan D3	$C_{25}H_{25}{}^2H_3N_6O$	8.8
69	Irgarol D9 (+)	$C_{11}H_{10}{}^2H_9N_5S$	9.8
70	Isoproturon D6 (+)	$C_{12}H_{12}{}^2H_6N_2O$	9.7
71	Lamotrigine 13C3,D3 (+)	$C_6{}^{13}C_3H_4{}^2H_3Cl_2N_5$	5.4
72	Levetiracetam D3 (+)	$C_8H_{11}{}^2H_3N_2O_2$	4.8
73	Lidocaine D10 (+)	$C_{14}H_{12}{}^2H_{10}N_2O$	5.3
74	Linuron D6	$C_9H_4{}^2H_6Cl_2N_2O_2$	11.4
75	MCPA D3 (-)	$C_9H_6{}^2H_3ClO_3$	9.8
76	Mecoprop D6 (-)	$C_{10}H_5{}^2H_6ClO_3$	10.6
77	Mefenamic acid D3	$C_{15}H_{12}{}^2H_3NO_2$	13.2
78	Mesotrione D3	$C_{14}H_{10}{}^2H_3NO_7S$	8.8
79	Metalaxyl D6 (+)	$C_{15}H_{15}{}^2H_6NO_4$	9.8
80	Methiocarb D3 (+)	$C_{11}H_{12}{}^2H_3NO_2S$	11.2
81	Methylprednisolone D3 (+)	$C_{22}H_{27}{}^2H_3O_5$	8.4
82	Metolachlor D6 (+)	$C_{15}H_{16}{}^2H_6ClNO_2$	12.8
83	Metolachlor-ESA D11	$C_{15}H_{12}{}^2H_{11}NO_5S$	7.2
84	Metoprolol D7 (+)	$C_{15}H_{18}{}^2H_7NO_3$	5.6
85	Metronidazole D4 (+)	$C_6H_5{}^2H_4N_3O_3$	4.7
86	Metsulfuron-methyl D3	$C_{14}H_{12}{}^2H_3N_5O_6S$	8.8
87	Morphine D3 (+)	$C_{17}H_{16}{}^2H_3NO_3$	4.3
88	N,N-Diethyl-3-methylbenzamide D10 (+)	$C_{12}H_7{}^2H_{10}NO$	9.8
89	N,O-Didesmethyl venlafaxine D3 (+)	$C_{15}H_{20}{}^2H_3NO_2$	5.1
90	N4-Acetyl-sulfathiazole D4	$C_{11}H_7{}^2H_4N_3O_3S_2$	5.4
91	Naproxen D3 (+)	$C_{14}H_{11}{}^2H_3O_3$	10.3
92	Nelfinavir D3	$C_{32}H_{42}{}^2H_3N_3O_4S$	8.9
93	Nicosulfuron D6	$C_{15}H_{12}{}^2H_6N_6O_6S$	7.8
94	Octhilinone D17 (+)	$C_{11}H_2{}^2H_{17}NOS$	11.5
95	O-Desmethylvenlafaxine D6 (+)	$C_{16}H_{19}{}^2H_6NO_2$	5.2
96	Oxazepam D5	$C_{15}H_6{}^2H_5ClN_2O_2$	8.8
97	Oxcarbazepine D4 (+)	$C_{15}H_8{}^2H_4N_2O_2$	7.5
98	Paracetamol D4 (+)	$C_8H_5{}^2H_4NO_2$	4.7
99	Phenazone D3 (+)	$C_{11}H_9{}^2H_3N_2O$	5.8
100	Pirimicarb D6 (+)	$C_{11}H_{12}{}^2H_6N_4O_2$	5.9
101	Pravastatin D3 (-)	$C_{23}H_{33}{}^2H_3O_7$	8.1
102	Primidone D5 (+)	$C_{12}H_9{}^2H_5N_2O_2$	5.8
103	Prochloraz D7 (+)	$C_{15}H_9{}^2H_7Cl_3N_3O_2$	11.0
104	Propamocarb free base D7 (+)	$C_9H_{13}{}^2H_7N_2O_2$	4.6
105	Propazine D6 (+)	$C_9H_{10}{}^2H_6ClN_5$	11.0
106	Propiconazole D5 (+)	$C_{15}H_{12}{}^2H_5Cl_2N_3O_2$	13.0
107	Propranolol D7 (+)	$C_{16}H_{14}{}^2H_7NO_2$	6.7
108	Pyrimethanil D5 (+)	$C_{12}H_8{}^2H_5N_3$	9.1
109	Ranitidine D6	$C_{13}H_{16}{}^2H_6N_4O_3S$	4.5
110	Ritalinic acid D10 (+)	$C_{13}H_7{}^2H_{10}NO_2$	5.2
111	Ritonavir D6 (+)	$C_{37}H_{42}{}^2H_6N_6O_5S_2$	12.4
112	Simazine D5 (+)	$C_7H_7{}^2H_5ClN_5$	8.3
113	Sotalol D6	$C_{12}H_{14}{}^2H_6N_2O_3S$	4.5
114	Sulcotrione D3	$C_{14}H_{10}{}^2H_3ClO_5S$	9.0
115	Sulfadiazine D4	$C_{10}H_6{}^2H_4N_4O_2S$	5.1
116	Sulfadimethoxine D4	$C_{12}H_{10}{}^2H_4N_4O_4S$	7.7
117	Sulfamethazine 13C6	$C_6{}^{13}C_6H_{14}N_4O_2S$	5.9
118	Sulfamethoxazole D4	$C_{10}H_7{}^2H_4N_3O_3S$	6.8
119	Sulfapyridine D4	$C_{11}H_7{}^2H_4N_3O_2S$	5.3
120	Sulfathiazole D4	$C_9H_5{}^2H_4N_3O_2S_2$	5.1
121	Tebuconazole D6 (+)	$C_{16}H_{16}{}^2H_6ClN_3O$	12.2
122	Terbuthylazine D5 (+)	$C_9H_{11}{}^2H_5ClN_5$	11.3
123	Terbutryn D5 (+)	$C_{10}H_{14}{}^2H_5N_5S$	9.4
124	Thiamethoxam D3 (+)	$C_8H_7{}^2H_3ClN_5O_3S$	5.7
125	Tramadol D6 (+)	$C_{16}H_{19}{}^2H_6NO_2$	5.6
126	Trimethoprim D9 (+)	$C_{14}H_9{}^2H_9N_4O_3$	4.9
127	Valsartan 13C5,15N	$C_{19}{}^{13}C_5H_{29}N_4{}^{15}NO_3$	10.8
128	Valsartan acid D4	$C_{14}H_6{}^2H_4N_4O_2$	7.3
129	Venlafaxine D6 (+)	$C_{17}H_{21}{}^2H_6NO_2$	6.3
130	Verapamil D6 (+)	$C_{27}H_{32}{}^2H_6N_2O_4$	8.1

(+): Compound only detectable in ESI positive mode

(-): Compound only detectable in ESI negative mode

Table S5. Parameters for data processing

Software package	Parameter	Value
	Peak finding	
	Minimum retention time	2 min
	Maximum retention time	19 min
	Subtraction offset	10 scans
	Subtraction multiple factor	1.3
	Noise threshold	100 cps
MarkerView™	Minimum spectral peak width	10 ppm
(1.2.1)	Minimum RT peak width	3 scans
	Assign charge states	enabled
	Peak alignment	
	Retention time tolerance	0.15 min
	Mass tolerance	20 ppm
	Intensity threshold	100 cps
	Peak integration	
	Integration algorithm	MQ4
	Gaussian smooth width	1.0 points
	Mass tolerance EIC	±10 ppm
MultiQuant™	Retention time half window	9 sec
(3.0.2)	Report largest peak	disabled
	Min. peak width	6 points
	Min peak height	100 cps
	Noise percentage	99.9%
	Baseline sub. window	0.3 min
	Peak splitting	2 points

Table S5. Parameters for data processing (continued)

Software package	Parameter	Value
	Filtering	
	Blank correction; all models	
	$f_A = \dfrac{\text{Peak Area}_{\text{Sample}}}{\text{Peak Area}_{\text{Blank}}}$ *)	$[\,5\,,\infty\,[$
	Absolute threshold; all models	
	Peak area (A)	$[\,350\,,\infty\,[$ cts
	Peak height (H)	$[\,100\,,\infty\,[$ cps
	Retention time (R)	$[\,2.0\,,19.0\,]$ min
	Peak width at 50% (W)	$[\,3.5\,,12.0\,]$ sec
	Ratio of peak area/peak height (A/H)	$[\,3.5\,,12.0\,]$ sec
	Relative threshold; model (b, c, d)	
MATLAB (R2015a)	Ratio peak area $(\frac{A_i}{A_j})$	$\left[\frac{2}{3}\,,\frac{3}{2}\right]$
	Ratio peak height $(\frac{H_i}{H_j})$	$\left[\frac{2}{3}\,,\frac{3}{2}\right]$
	Ratio retention time $(\frac{R_i}{R_j})$	$\left[\frac{99}{100}\,,\frac{100}{99}\right]$
	Ratio peak width at 50% $(\frac{W_i}{W_j})$	$\left[\frac{2}{3}\,,\frac{3}{2}\right]$
	Ratio peak area/peak height $(\frac{A/H_i}{A/H_j})$	$\left[\frac{2}{3}\,,\frac{3}{2}\right]$
	Relative threshold $Q_{n,k}$ using different injection levels; model (e) only *Here: Level 1 = 50 µL, Level 2 = 75 µL and Level 3 = 100 µL*	
	$Q_{1,2} = \dfrac{\text{Area}_1 \cdot \text{Volume}_2}{\text{Area}_2 \cdot \text{Volume}_1}$ **)	$\left[\frac{9}{14}\,;\frac{3}{2}\right]$
	$Q_{1,3} = \dfrac{\text{Area}_1 \cdot \text{Volume}_3}{\text{Area}_3 \cdot \text{Volume}_1}$ **)	$\left[\frac{2}{3}\,;\frac{10}{7}\right]$
	$Q_{2,3} = \dfrac{\text{Area}_2 \cdot \text{Volume}_3}{\text{Area}_3 \cdot \text{Volume}_2}$ **)	$\left[\frac{20}{27}\,;\frac{4}{3}\right]$

*). The same factor f_A is also applied to proof whether the peak height differs f_a-times

**).The same $Q_{n,k}$ are applied to proof whether the peak height is within the tolerance

(a)

(b)
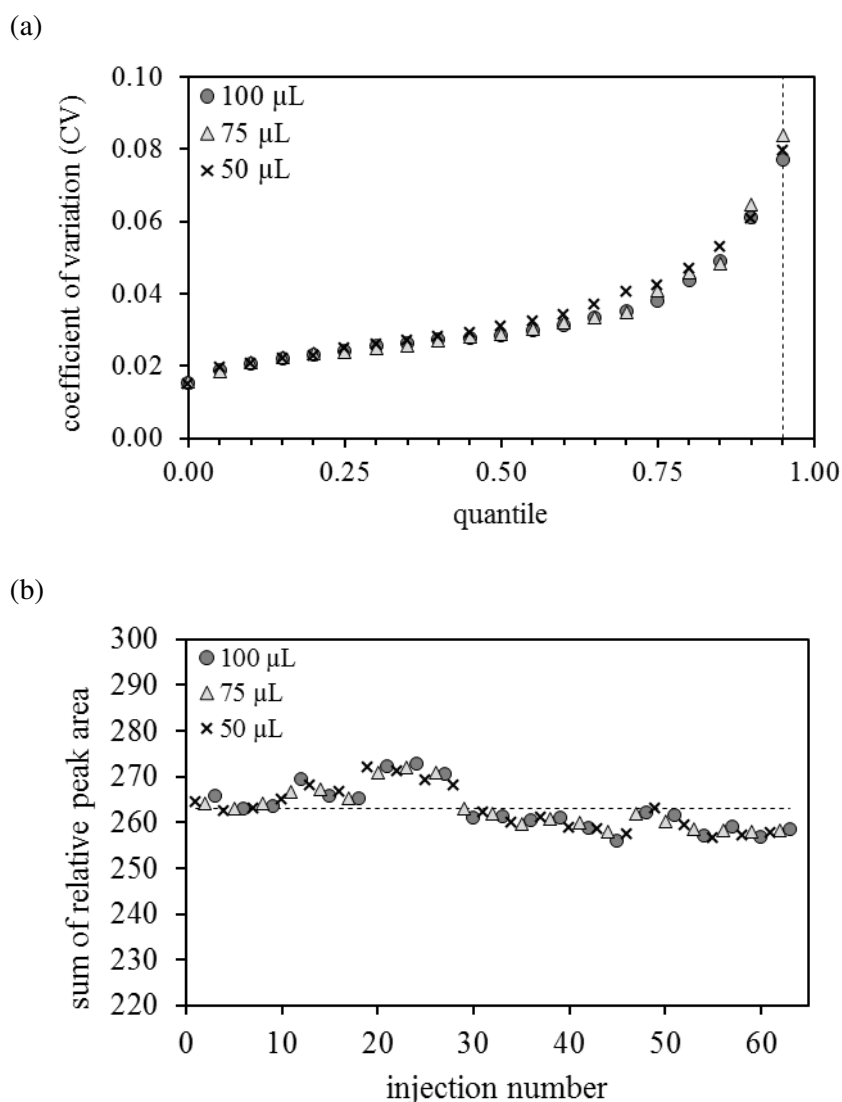
Figure S1. Stability tests throughout the entire period of analysis;
(a) coefficient of variation (CV) as a function of the quantile; CVs were calculated from 263 standards (*Table S2*) at three different injection volumes (50, 75 and 100 µL) from a measurement series of 63 replicates (hence, 21 injections at each injection level), the dashed line (---) represents the 95%-quantile; (b) sum of the relative peak area of all 263 standards against the consecutive number of injections with alternating injection volumes (50, 75, 100, 50, …), the relative peak areas are based on the mean value of the respective standard at a given injection volume; the dashed line (---) indicates the expected value of N = 263.
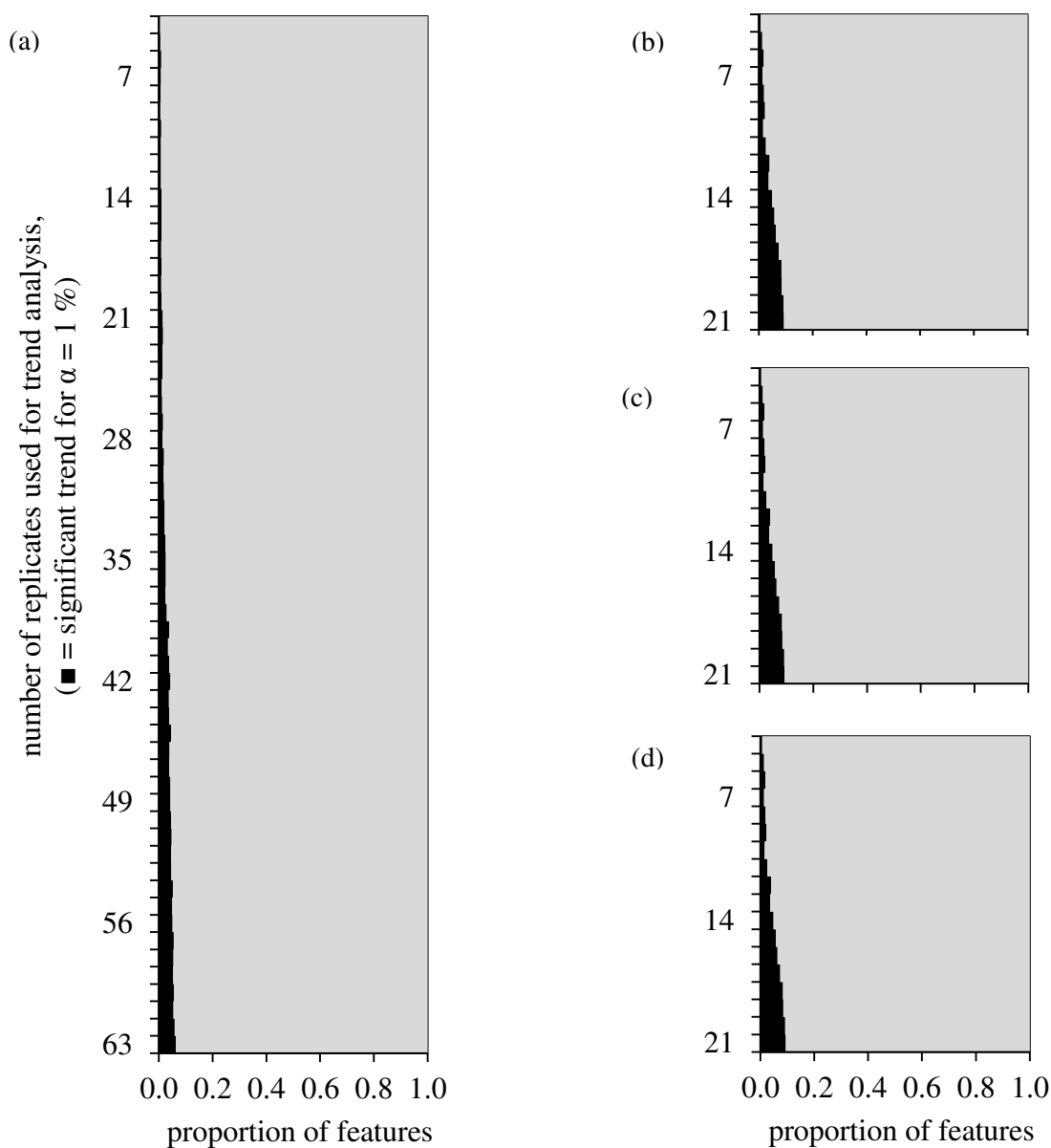
Figure S2. Trend analysis according to von Neumann;
(a) dataset of 63 zero-injections (system blank, $N_{feature}$ = 2295) over a total analysis time of more than 37 h, dataset of sample -A- ($N_{feature}$ = 7337) at (b) 50 µL, (c) 75 µL and (d) 100 µL injection volume over a total analysis time of almost 40 h; the ordinate shows the number of replicates that were chosen to form the sample while on the abscissa the proportion of features showing a statistically significant trend (■ = percentage of features with test values higher than the critical limit for p = 99%) is illustrated;

Note: The approximation function $CL = 2 - 2 \cdot z_{1-\alpha}\sqrt{\frac{n-2}{n^2-1}}$ *was used to estimate the critical limit (CL) for n > 60 (negligible deviations for n > 50 between estimated and tabulated values, maximal deviation was less than 0.6%)*

Explanation S1. Converting the number of combinations (C) into the number of remaining injections (r)

Assuming a particular feature would occur in all n technical replicates and, in addition, also passing through all of the specified thresholds, the algorithm would provide the number of all theoretically possible combinations C_{max}. Conversely, knowing that a feature has reached C_{max}, it can be concluded that it occurred exactly n times. However, some features do only partly fulfill the threshold criteria, hence leading to a value smaller than C_{max}. However, in reality some features do only partly fulfill the different threshold criteria, leading to combination values C which are smaller than C_{max}. A frequent example where this situation arises represents features whose intensities lie very close to the threshold. In some cases the actual intensity is slightly above, in other cases slightly beyond the specified threshold. For this reason, only combinations which include samples with features above the threshold will contribute to C. Consequently, the number of all replicates (n) is reduced to the number of the remaining replicates ($r = n$ - x) with x being the number of replicates that did not pass the thresholds. From the combinatorial perspective, a function $f(r) = C$ can be derived for model (a) to (d) which maps each number of remaining injections (r) to the corresponding combination value (C). The inverse function $f^{-1}(C) = r$, on the other hand, would allow to map each combination value to the number of remaining replicates. The derivation of the inverse function, however, becomes complicated for higher-degree polynomial functions which is why the function $f^{-1}(C) = r$ was numerically solved.

As a first step, the $\binom{n}{k}$-functions were written in their polynomial forms which were numerically solved using "Root of nonlinear function" (Matlab) for all integer C numbers ($C \in \mathbb{N}$) in the codomain of interest. This allows mapping each number of combinations (C) back to the number of remaining injections (r) without the formation of the actual inverse function (within the domain of definition of these functions, each C-value can be assigned to one unique r-value). Note that the number of achievable combinations (count cases that fulfill the filters) must be positive integer numbers ($C \in \mathbb{N}$) while the number of remaining injections - after conversion - can also accept non-integer values ($r \in \mathbb{R}^+$).

Model (e), however, demonstrates a special case as, depending on the distribution of the replicates that do not pass the thresholds to the three injection levels, multiple numbers of combinations are obtained for the same number of remaining replicates (r). By way of example, the dataset in this study consists of 63 injections (21 at each level). Assuming features beyond the tolerance in 18 injections it seems evident, that the case where the outliers are uniformly distributed over three injection levels (hence six injections per level) results in a higher number of combinations as if the outliers exclusively occur at one level. Taking into account the outlier distribution, the general formula to calculate the number of possible combinations is composed as follows:

$$C = (n - x)(n - y)(n - z)$$

with x, y and z being the number of outliers (i.e. features that do not pass the threshold criteria) at the respective injection level. The remaining number of injections r can be expressed by $r = 3n - (x + y + z)$ and after transformation of $C(n)$ to $C(r)$ the equation describes the number of achievable combinations for different outlier distributions.

First of all, the combinations C were calculated for all feasible constellations of x, y and z before the following limit cases were derived:

(I)   *x = y = z*
(II)   *x = y; z = x - 1*
(III)   *y = z = x - 1*
(IV)   *y = z = 0*
(V)   *y = (n - 1); z = 0*
(VI)   *y = z = (n - 1)*

As shown in <u>*Figure S3*</u>, the different functions derived from (I) to (VI) describe the extreme values (black) while the possible combinations for all outlier distributions (grey) are within these limits.
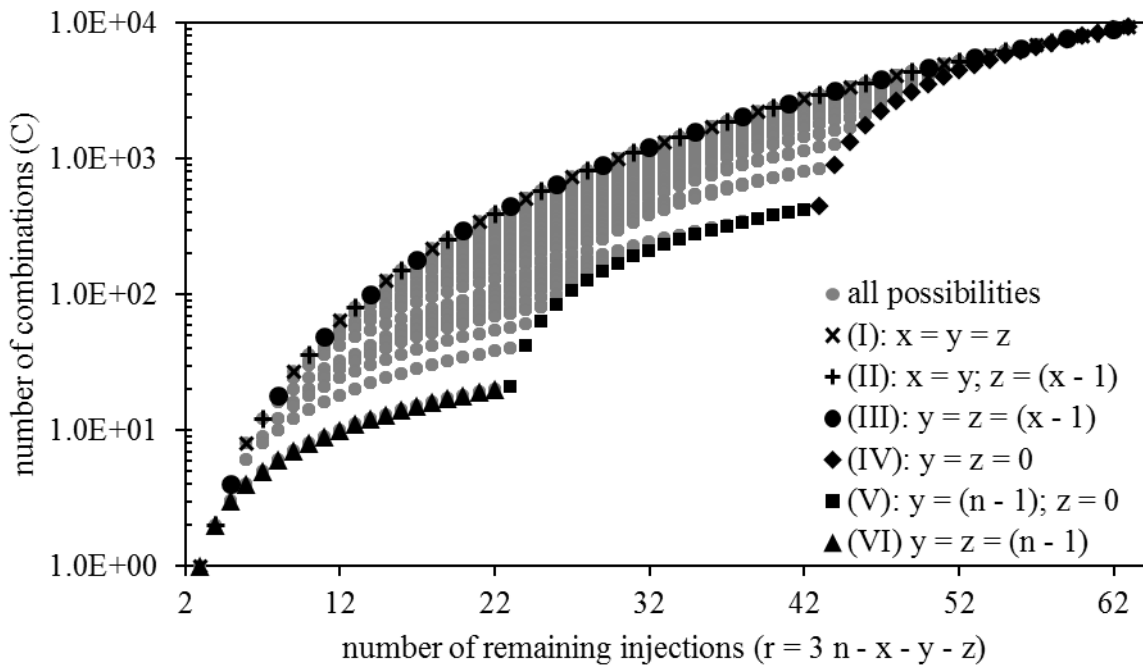
Figure S3. Simulated combination values for all possible outlier distributions

To take but one example, 20 outliers at injection level 3 ($z = 20$) would result in a remaining number of $r = 63 - (0 + 0 + 20) = 43$ injections and in a combination value of $C = 21 \cdot 21 \cdot (21 - 20) = 441$. A more uniform distribution of the 20 outliers such as $x = 6$, $y = 7$ and $z = 7$ would also lead to a remaining number of replicates $r = 63 - (6 + 7 + 7) = 43$ but to a combination value of $C = (21 - 6) \cdot (21 - 7) \cdot (21 - 7) = 2940$. As evident, the distribution of the injections that fail the threshold criteria to the three injection level has a high impact on the achievable combination value. Conversely, having one combination value C, multiple constellations of x, y and z and therefore r are possible.

After processing the data using model (e), the algorithm delivers the combination values for each single feature but not the distribution of the outliers to the three levels. Consequently, a clear assignment of the remaining injections (r) to the obtained combination value (C) is no longer possible. Considering the extreme values, hence the best- and worst-case scenarios, at least a possible range of the r-values can be localized. The deviation between both scenarios can be taken from *Figure S4* which comprises all possible constellations. The maximum deviation of about 20 replicates is reached for 43 remaining injections (x=43.00, y= 22.83). As can be seen, for higher values, on the other hand, the deviations become very low and thus leading to substantially smaller differences.

(a)   (b)
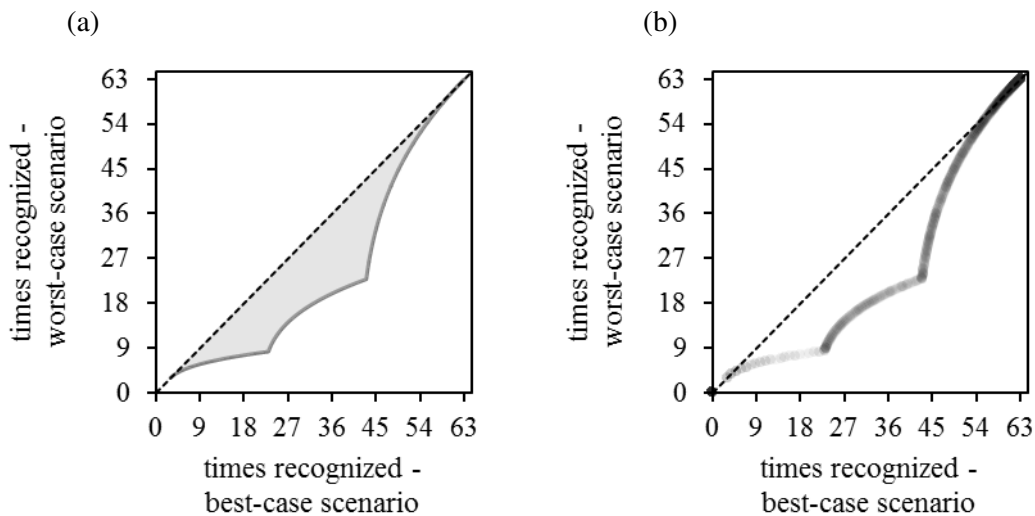
Figure S4. Deviations in the remaining injection numbers;
Deviation between the best- and worst-case scenario caused by conversion from the reached combinations (C) back to the remaining number of injections (r), (a) shows the possible deviations based on the theory whereas (b) illustrates the results obtained from a real data set (sample -A-), *Note: In case of (b), the density of the point (shading) also shows the distribution*

In overall terms, the deviations between both scenarios could be very small if most features possess high combination values and hence high r-values. For instance, the evaluation of sample -A- (see main part) is depicted in *Figure S4 (b)* where the distribution of the r-values is indicated by the point density (shading) which reaches the highest values for very high r-values (times recognized >50). In these cases, the differences between both extreme cases remain very low with the result that the calculated mean values could be closer than expected.

All equations used in this study are plotted in *Figure S5* and the exact equations which were derived from the combinational approach are shown below.
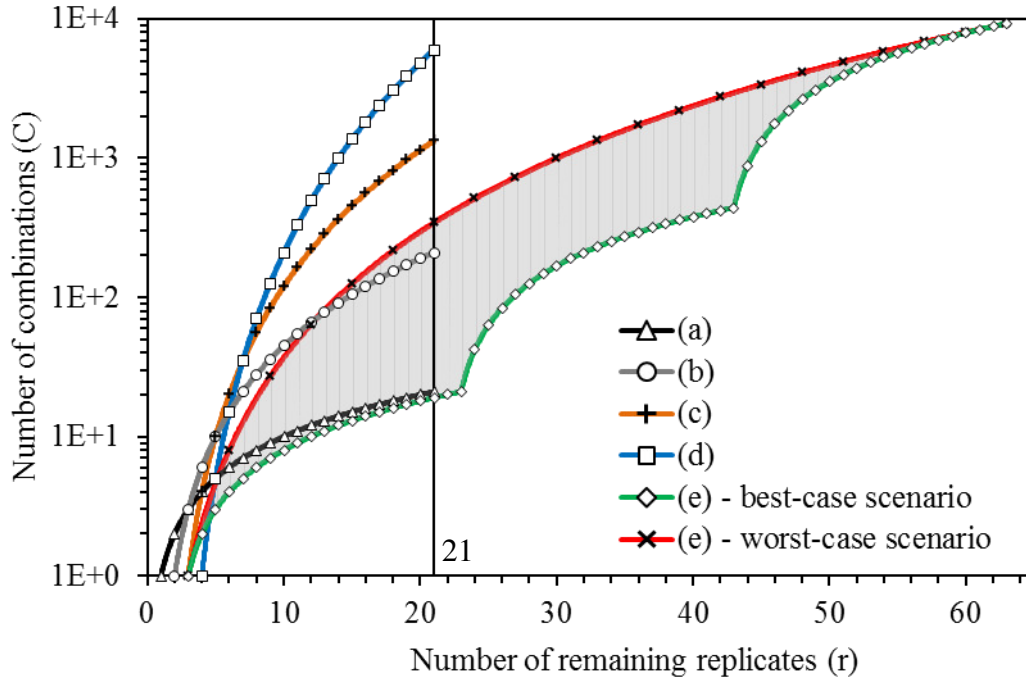
Figure S5. Number of combinations (*C*) as functions of the injection numbers (*r*); (a) single replicates, (b) replicates duplicates, (c) replicates triplicates, (d) replicates quadruplicates and (e) replicates triplicates at three different injection levels;
Note: In case of (e) 21 times three injection volumes (here: 50, 75 and 100 µL) were evaluated adding to a total number of 63 injections.

$$C_a(n) = n \qquad\qquad for \quad 1 \leq n \leq 21$$

$$C_b(n) = \frac{1}{2}n^2 - \frac{1}{2}n \qquad\qquad for \quad 2 \leq n \leq 21$$

$$C_c(n) = \frac{1}{6}n^3 - \frac{1}{2}n^2 + \frac{1}{3}n \qquad\qquad for \quad 3 \leq n \leq 21$$

$$C_d(n) = \frac{1}{24}n^4 - \frac{1}{4}n^3 + \frac{11}{24}n^2 - \frac{1}{4}n \qquad for \quad 4 \leq n \leq 21$$

$$C_{e,worst}(n) = \frac{1}{27}n^3 \qquad\qquad for \quad 3 \leq n \leq 63$$

$$C_{e,best}(n) = \begin{cases} 441n - 18522 & for \quad 43 \leq n \leq 63 \\ 21n - 462 & for \quad 23 \leq n < 43 \\ n - 2 & for \quad 3 \leq n < 23 \end{cases}$$
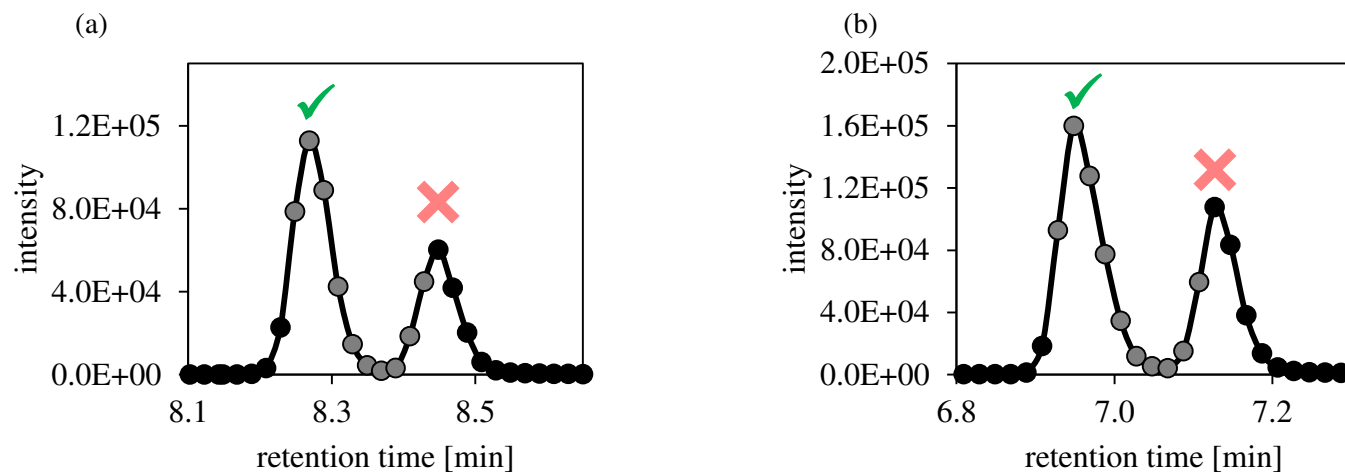
Figure S6. Problem cases using background subtraction;
background subtraction (10 scans offset, grey highlighted) is leading to false negative results for the later eluting isomers; (a) Simazine ($C_7H_{12}ClN_5$; RT ≈ 8.25), Desethylterbutylazin ($C_7H_{12}ClN_5$; RT ≈ 8.45); (b) Desmetryn ($C_8H_{15}N_5S$; RT ≈ 6.95), Irgarol-descyclopropyl ($C_8H_{15}N_5S$; RT ≈ 7.15);
Note: The background subtraction is accomplished by subtraction of the MS spectrum 10 scans before (grey points) the peak apex which was multiplied by a certain subtraction multiple factor (here: 1.3)

19

(a)
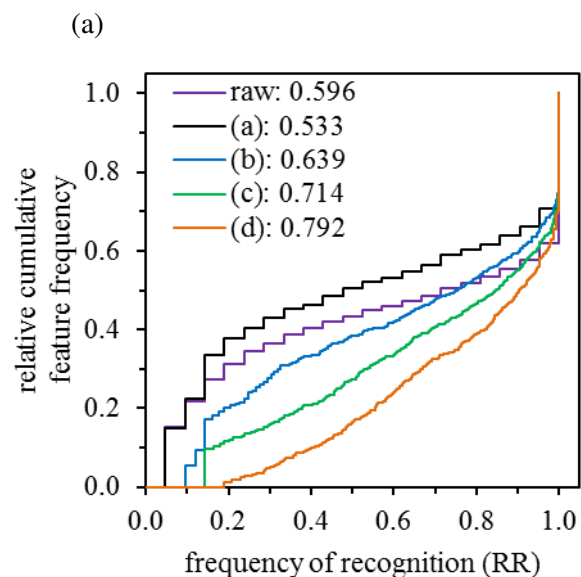
(b)

(c)

(d)
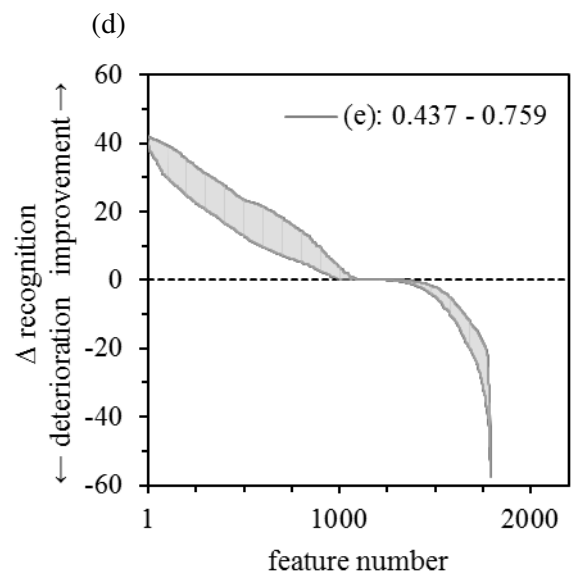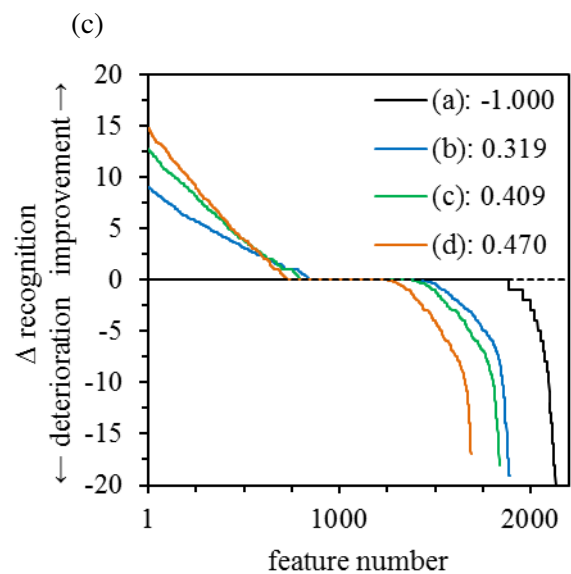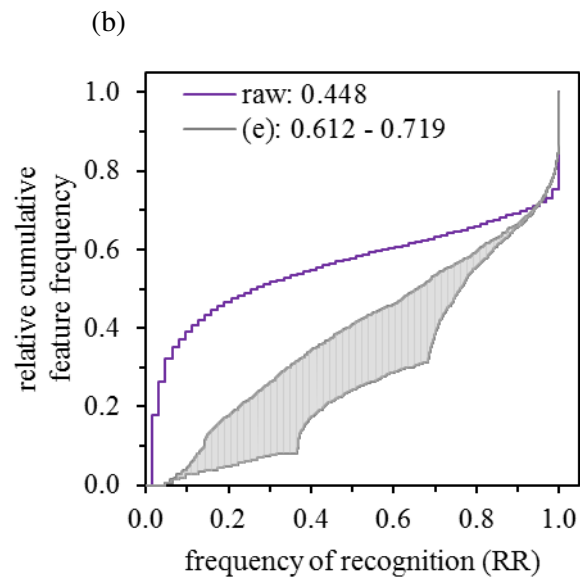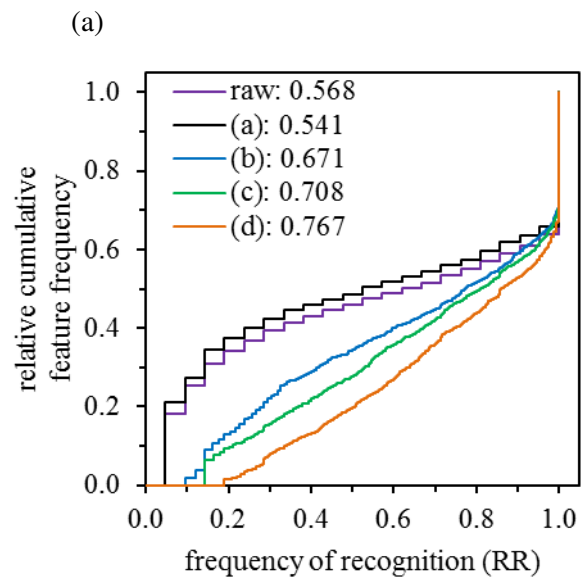
Figure S7. Comprehensive evaluation of sample -B- (ultrapure water + 1000 ng L^{-1} of stock solution I)

(a) Cumulative feature distribution of the raw data and model (a)-(d), 21 injections, 100 µL each

(b) Cumulative feature distribution of the raw data and model (e), 63 injection (50, 75, 100 µL)

Note: The mean rate of recognitions are depicted in the legend behind the respective entry

(c) Δ recognition values plotted in descending order for all features, model (a) - (d) with respect to the raw data, 21 injections, 100 µL each

(d) Δ recognition values plotted in descending order for all features, model (e) with respect to the raw data, 63 injection (50, 75, 100 µL)

Note: The mean improvement factors are depicted in the legend behind the respective entry

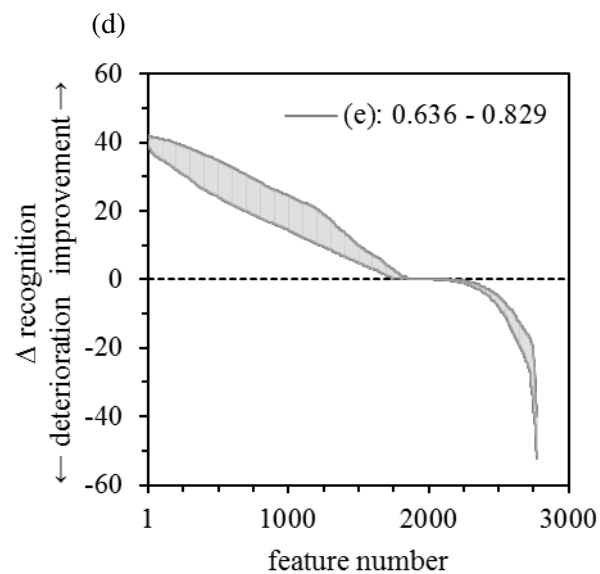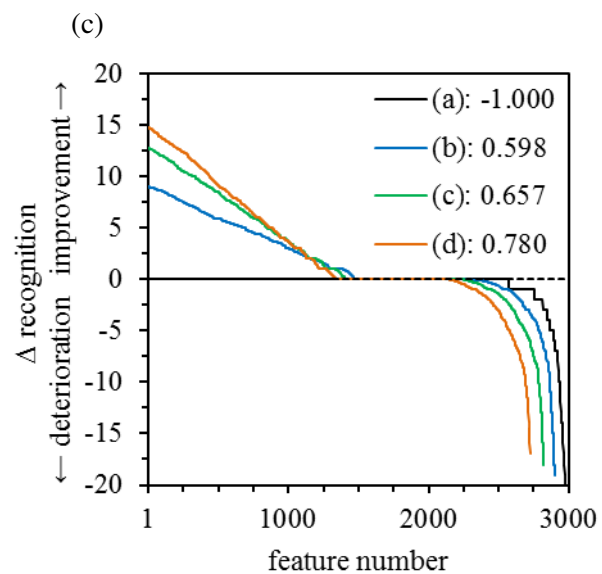*Note: For model (e) the best- and worst-case scenarios were taken into consideration*
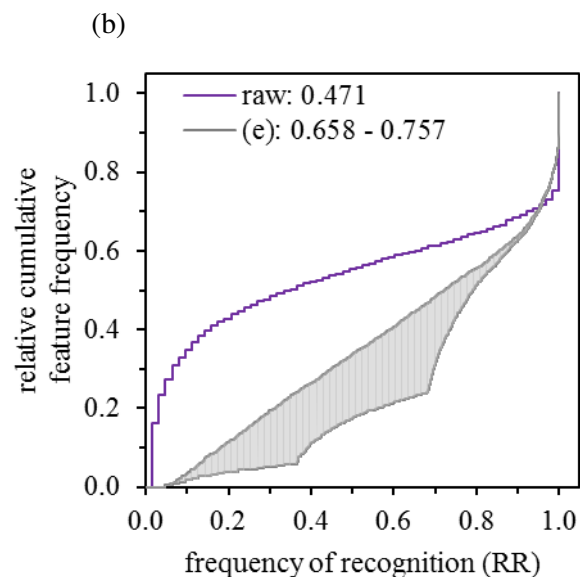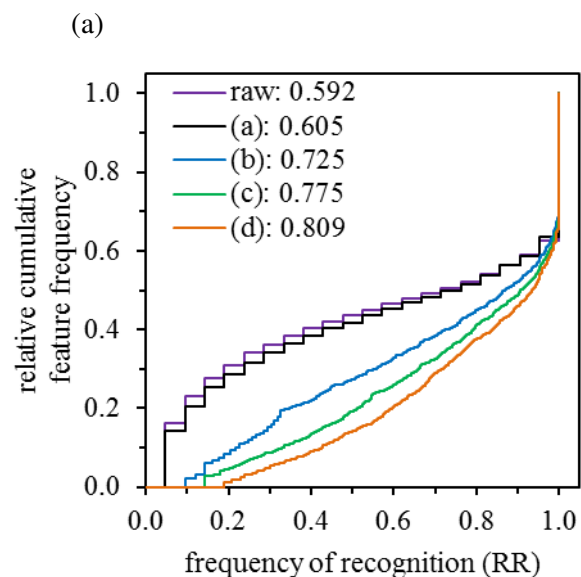
20

(a)

(b)

(c)

(d)

Figure S8. Comprehensive evaluation of sample -C- (River water + 100 ng L^{-1} of stock solution I) (a) Cumulative feature distribution of the raw data and model (a)-(d), 21 injections, 100 µL each

(b) Cumulative feature distribution of the raw data and model (e), 63 injection (50, 75, 100 µL)

Note: The mean rate of recognitions are depicted in the legend behind the respective entry

(c) Δ recognition values plotted in descending order for all features, model (a) - (d) with respect to the raw data, 21 injections, 100 µL each

(d) Δ recognition values plotted in descending order for all features, model (e) with respect to the raw data, 63 injection (50, 75, 100 µL)

Note: The mean improvement factors are depicted in the legend behind the respective entry

Note: For model (e) the best- and worst-case scenarios were taken into consideration

21

Figure S9. Comprehensive evaluation of sample -D- (Stagnated tap water + 100 ng L^{-1} of stock solution I)

(a) Cumulative feature distribution of the raw data and model (a)-(d), 21 injections, 100 µL each

(b) Cumulative feature distribution of the raw data and model (e), 63 injection (50, 75, 100 µL)

Note: The mean rate of recognitions are depicted in the legend behind the respective entry

(c) Δ recognition values plotted in descending order for all features, model (a) - (d) with respect to the raw data, 21 injections, 100 µL each

(d) Δ recognition values plotted in descending order for all features, model (e) with respect to the raw data, 63 injection (50, 75, 100 µL)

Note: The mean improvement factors are depicted in the legend behind the respective entry

Note: For model (e) the best- and worst-case scenarios were taken into consideration
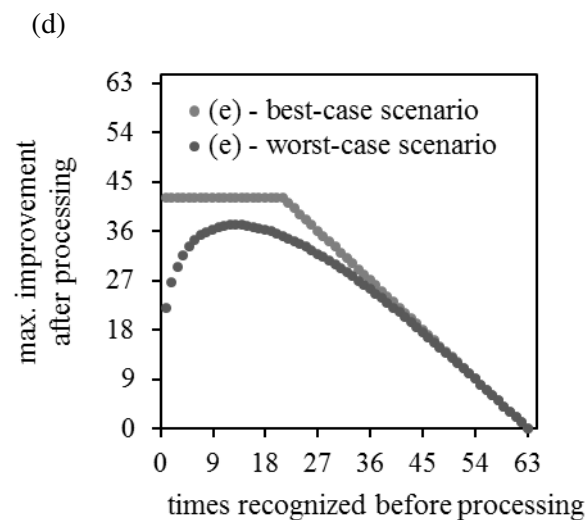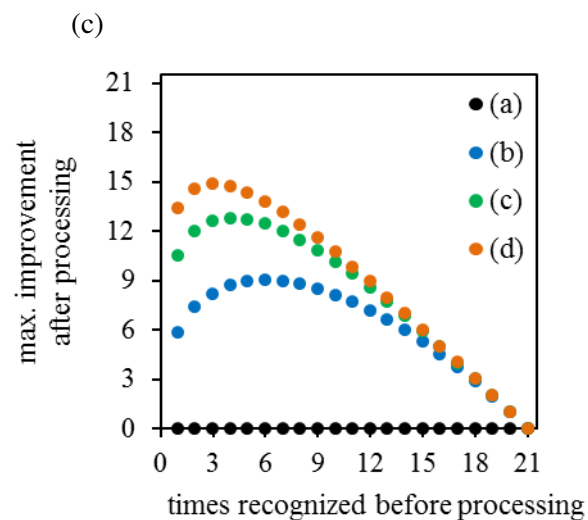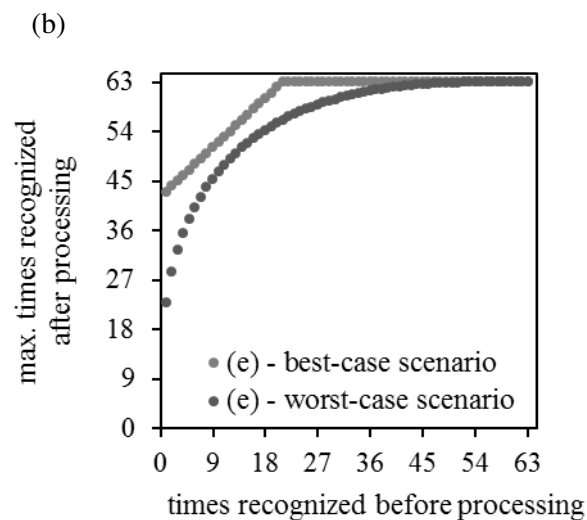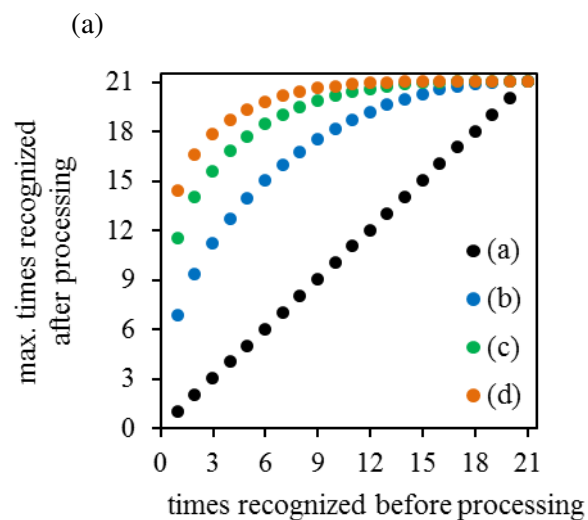
22

(a)

(b)

(c)

(d)

Figure S10. Theoretical calculations to derive the maximum achievable extent of improvement for the respective model

(a) x-y-scatterplot for model (a) to (d) showing the maximum achievable number that could theoretically be obtained by the respective model (combinatorial approach)

(b) x-y-scatterplot for model (e) showing the maximum achievable number that could theoretically be obtained, the best- and worst-case scenario was simulated

(c) Maximum achievable improvement with respect to the raw data for model (a) to (d)

(d) Maximum achievable improvement with respect to the raw data for model (e)), the best- and worst-case scenario was simulated

Note: The maximum improvement was simulated for cases where, for example, a feature was only recognized once, although a true peak was apparent in each replicate (false negative). Therefore, each combination containing the specific replicate would result in a positive hit

23

Table S6. Improvement or deterioration with respect to the raw data (normal and weighted with the extent)

	$\sum$	Number of features (relative part)			Weighted number of features		
		improvement	no change	deterioration	Area (+)	Area (-)	$\bar{I}$
Sample -A-							
Model (a)	4441	0 (0)	4148 (0.934)	293 (0.066)	0	1578	-1.000
Model (b)	4369	1838 (0.421)	1816 (0.416)	715 (0.164)	8949	1783	0.668
Model (c)	4298	1771 (0.412)	1811 (0.421)	716 (0.167)	12054	1933	0.724
Model (d)	4215	1716 (0.407)	1785 (0.423)	714 (0.169)	13279	2001	0.738
Model (e)-best	4301	2248 (0.523)	864 (0.201)	1189 (0.276)	38263	7817	0.661
Model (e)-worst	4301	2340 (0.544)	862 (0.2)	1099 (0.256)	55529	5303	0.826
Sample -B-							
Model (a)	2135	0 (0)	1884 (0.882)	251 (0.118)	0	1296	-1.000
Model (b)	1891	836 (0.442)	583 (0.308)	472 (0.25)	3487	1800	0.319
Model (c)	1841	797 (0.433)	583 (0.317)	461 (0.25)	4629	1941	0.409
Model (d)	1688	724 (0.429)	531 (0.315)	433 (0.257)	5010	1805	0.470
Model (e)-best	1791	996 (0.556)	239 (0.133)	556 (0.31)	14787	5793	0.437
Model (e)-worst	1791	1095 (0.611)	225 (0.126)	471 (0.263)	24439	3341	0.759
Sample -C-							
Model (a)	844	0 (0)	733 (0.868)	111 (0.132)	0	661	-1.000
Model (b)	815	411 (0.504)	212 (0.26)	192 (0.236)	1669	695	0.412
Model (c)	747	379 (0.507)	203 (0.272)	165 (0.221)	2164	570	0.583
Model (d)	667	342 (0.513)	168 (0.252)	157 (0.235)	2324	475	0.661
Model (e)-best	743	494 (0.665)	52 (0.07)	197 (0.265)	7559	1653	0.641
Model (e)-worst	743	543 (0.731)	49 (0.066)	151 (0.203)	11731	1088	0.830
Sample -D-							
Model (a)	2978	0 (0)	2567 (0.862)	411 (0.138)	0	1624	-1.000
Model (b)	2899	1470 (0.507)	839 (0.289)	590 (0.204)	6794	1706	0.598
Model (c)	2822	1413 (0.501)	835 (0.296)	574 (0.203)	9071	1880	0.657
Model (d)	2725	1341 (0.492)	801 (0.294)	583 (0.214)	9926	1952	0.671
Model (e)-best	2774	1747 (0.63)	255 (0.092)	772 (0.278)	30134	6694	0.636
Model (e)-worst	2774	1856 (0.669)	245 (0.088)	673 (0.243)	44870	4208	0.829

Explanation S2. Context between intensity threshold and peak height or peak area

Unfortunately, the exact way of how the peak finding algorithm works is not known (corporate secret), therefore leading to uncertainties if comparing manually processed data with data obtained from the algorithm. One fundamental parameter in all evaluations is the intensity threshold that determines which signals are processed (*signal > threshold*) and which ones remain disregarded (*signal ≤ threshold*). The intensity threshold possesses the dimension counts per second (cps) whereas the generated response shows another one. While we could show that in general a good correlation between the reported response and the real peak area and peak height exist (*Figure S11*), we have been struggling to find the exact relationship which is indispensable to elucidate the connection between the intensity threshold and the peak area or peak height. Once having found a possible peak location, the algorithm sums the spectra from the starting to the ending scan (i.e. across the LC peak) and then generates the response from the spectral peak area. This coincides with the fact that the peak area shows a slightly better correlation with the response. However, due to the fact that the limits (i.e. starting and ending scans) are automatically set by the algorithm, it is not possible to verify these calculated response values. For this reason it is not possible to directly compare the generated responses with the intensity threshold and moreover to exactly compare results to manually verified ones.
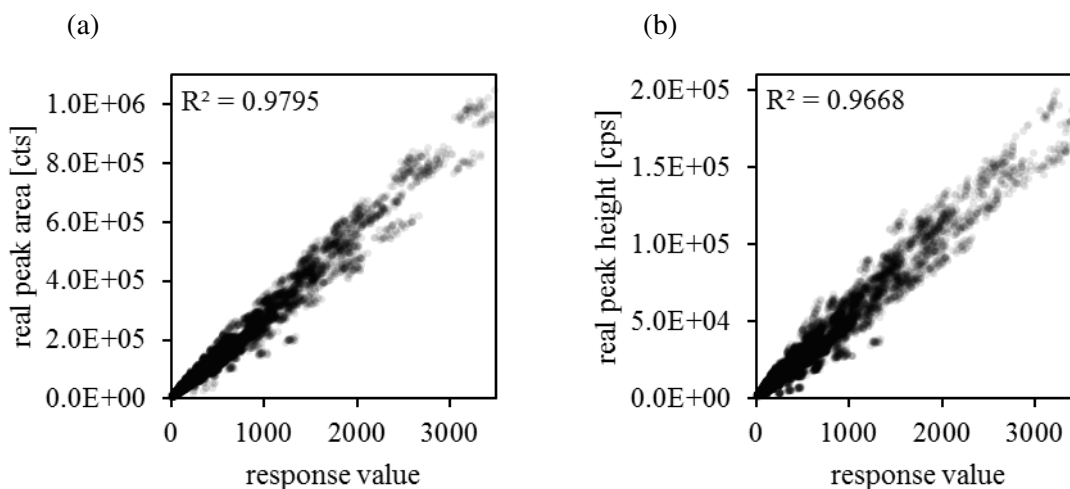
(a) (b)

Figure S11. Correlation between response and real peak area (a) and peak height (b)

PAPER 2

LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes

Tobias Bader, Wolfgang Schulz, Klaus Kümmerer, Rudi Winzenbacher

LC-HRMS Data Processing Strategy for Reliable Sample Comparison Exemplified by the Assessment of Water Treatment Processes

Tobias Bader,*,[†,‡] Wolfgang Schulz,[†] Klaus Kümmerer,[‡] and Rudi Winzenbacher[†]

[†]Laboratory for Operation Control and Research, Zweckverband Landeswasserversorgung, Am Spitzigen Berg 1, 89129 Langenau, Germany

[‡]Sustainable Chemistry and Material Resources, Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg, Scharnhorststraße 1/C13, 21335 Lüneburg, Germany

S *Supporting Information*

ABSTRACT: The behavior of micropollutants in water treatment is an important aspect in terms of water quality. Nontarget screening by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) offers the opportunity to comprehensively assess water treatment processes by comparing the signal heights of all detectable compounds before and after treatment. Without preselection of known target compounds, all accessible information is used to describe changes across processes and thus serves as a measure for the treatment efficiency. In this study, we introduce a novel LC-HRMS data processing strategy for the reliable classification of signals based on the observed fold changes. An approach for filtering detected features was developed and, after parameter adjustment, validated for its recall and precision. As proof of concept, the fate of 411 target compounds in a 0.1 μg/L standard mix was tracked throughout the data processing stages, where 406 targets were successfully recognized and retained during filtering. Potential pitfalls in signal classification were addressed. We found the recursive peak integration to be a key point for the reliable classification of signal changes across a process. For evaluating the repeatability, a combinatorial approach was conducted to verify the consistency of the final outcome using technical replicates of influent and effluent samples taken from an ozonation process during drinking water treatment. The results showed sufficient repeatability and thus emphasized the applicability of nontarget screening for the assessment of water treatment processes. The developed data processing strategies may be transferred to other research fields where sample comparisons are conducted.
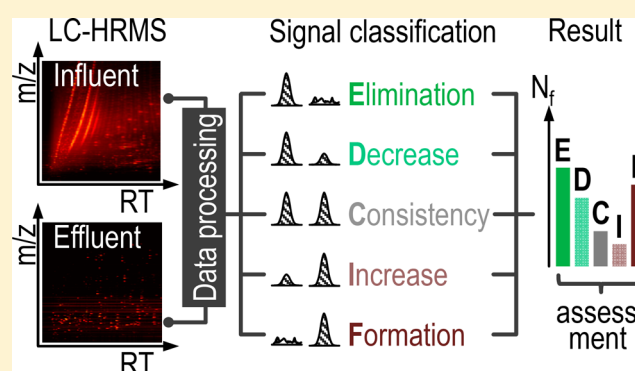
A variety of tens of thousands of organic substances are in daily use and manufactured, consumed, and disposed of in households or industry. Many of them, such as pharmaceuticals, pesticides, personal care products, and industrial chemicals, among others, have the potential to enter the aquatic environment as micropollutants.[1−4] Various targeted, as well as untargeted, monitoring studies revealed the presence of micropollutants in wastewater,[5−7] surface water,[8,9] and drinking water.[10−13]

Different treatment processes, for example, based on oxidation with ozone and/or adsorption onto activated carbon, for both wastewater and drinking water treatment have been developed to improve the removal of micropollutants.[14−17] To assess the performance of these treatment processes, removal efficiencies of known target compounds are generally quantified by liquid chromatography tandem mass spectrometry (LC-MS²). By such approaches, however, the complete evaluation of a process is based on a small subset of trace substances.[18,19] Assuming many transformation products (TPs) to be unknown yet, this group is particularly underrepresented by common strategies.[20] This restriction exposes the weakness of targeted

approaches and emphasizes the demand for new strategies which reveal a more comprehensive view of a treatment process.

Recent advances in high-resolution mass spectrometry (HRMS) coupled to liquid chromatography (LC) have initiated new possibilities for the detection of micropollutants at environmentally relevant concentrations in complex matrices.[21,22] Nontarget screening based on LC-HRMS becomes increasingly important as the information on all detectable compounds, comprising unknowns, is considered.[23−25] Many studies using nontarget screening were aimed at finding and identifying unknown or unexpected compounds, mostly transformation products.[26−30] Typically hundreds to thousands of chromatographic peaks (also called features) are detectable in each sample, making manual peak review and correction no longer a reasonable option. Instead, sophisticated processing strategies are needed to automatically process the wealth of

data, while keeping false positive and false negative findings to a minimum.[31−34]

The LC-HRMS data can also be used for the assessment of water treatment processes.[35] Even if most of the detected signals remain unknown, it is possible to track these features through treatment processes. The efficiency of the treatment can be evaluated by comparing the heights of the detected signals in the influent and effluent of the process, whereas care must be taken to ensure that matrix effects do not distort the relations. Thus, nontarget screening provides a more holistic picture of the entire process with less bias caused by preselection of known substances. However, nontarget screening does initially not provide reliable figures on quantitative concentrations (regulatory compliance).

Only few studies were published using HRMS as an evaluation tool to assess treatment processes. Merel et al. investigated wastewater samples treated with UV or UV/H$_2$O$_2$ in pilot scale treatment.[36] Nürenberg et al. classified features to assess the denitrification/nitrification step during biological wastewater treatment.[37] Parry et al. compared targeted and nontargeted approaches to evaluate the wastewater treatment by the advanced oxidation reactor.[38] Bader et al. assessed the fourth treatment step using granulated activated carbon after the biological wastewater treatment.[39] Heuett applied nontarget screening to investigate changes in the feature pattern across the process of drinking water treatment.[40] However, none of these studies addressed the reliability of the results which is a critical point in each untargeted workflow[41] and might affect the assessment of processes.

Therefore, the main goal of this work was to develop a strategy for assessing water treatment processes based on LC-HRMS data and verify the consistency of the final outcome. First, we developed a workflow for the reliable processing of LC-HRMS data, which was evaluated for its recall and precision.[42] Second, a concept for the assessment of water treatment processes based on the classification of (un)known signals was introduced. Potential pitfalls regarding the fold change (i.e., quotient of effluent and influent peak heights) calculation for signals close to the intensity threshold were discussed and an algorithm to overcome these difficulties was presented. Last, a combinatorial approach based on the comparison of multiple replicates was conducted to evaluate the repeatability of the complete workflow.

■ METHODS

Sampling. A grab sample taken from the River Danube at River Kilometer Index 2568 (Leipheim, Germany) was analyzed and served as training data for optimizing the extracted ion chromatogram (EIC) filtering. The test data set consisted of four grab samples taken from different stages of the drinking water treatment (Waterworks Langenau, Germany). The influent represented river water which was directly abstracted (no bank filtration) from the River Danube. Additional samples were taken from the effluents of ozonation (O$_3$), multilayer filtration (MLF) and activated carbon filtration (ACF). As further criterion for evaluating the workflow, a multicomponent standard (ultrapure water), comprising 411 target compounds (Supporting Information (SI), Tables S-1 and S-2) at 0.1 μg/L, was considered throughout the different stages of data processing. The targets belong to various substance classes, such as pesticides, pharmaceuticals, industrial chemicals and personal care products, and cover a broad mass (m/z 100−838) and polarity range (log D (pH 3) ≈ −1.5−5.0,

pH 3 because of acidified eluents). For assessing the repeatability of the developed strategy, the ozonation process in drinking water treatment was considered. The influent consisted of pretreated river water after flocculation and sedimentation which will hereinafter be referred to as river water. The effluent was taken after the process of ozonation. Both the influent and effluent samples were analyzed in nine technical replicates (repeated measurements of the same sample). To avoid discriminating against certain substances, no sample preparation was performed. The glassware used was treated for 4 h at 450 °C in a fusing oven (FE 230, Rohde, Germany) to prevent contamination of single samples.

Sample Acquisition. The LC-HRMS acquisition method is described in Table S-3. Briefly, samples were analyzed in triplicates using reversed phase separation (Zorbax Eclipse Plus C18, 2.1 × 150 mm, 3.5 μm, Agilent) coupled to a quadrupole-time-of-flight mass analyzer (TripleTOF 5600, Sciex). Ions were generated by positive electrospray ionization and monitored within m/z 100 and 1200. For quality assurance, 13 isotope-labeled internal standards (Table S-4), well distributed across the considered retention time range, were used. Samples and internal standards were coinjected using 95 and 5 μL, respectively. To obtain similar ion abundances for the internal standards (IS), different concentrations were used to prepare the spike solution. For blank correction, 95 μL of ultrapure water was coinjected with 5 μL of internal standards.

Software. MarkerView (1.2.1) was used for peak finding and peak alignment across different samples and replicates. The quantitation package MultiQuant (2.1.2 + Research Features, January 24, 2017) was utilized for subsequent peak integration of all detected features using two integration algorithms (*MQ4* and *Summation*). Both software tools are available from Sciex (Framingham, USA). Further processing steps were accomplished by in-house scripts executed by Matlab (R2016b, The MathWorks, Inc., Natick, USA).

Data Processing Workflow. The entire workflow was composed as follows: (1) peak finding, (2) peak alignment, (3) peak integration, and (4) EIC filtering. For peak finding and alignment, the specific parameters were optimized to keep the number of false negatives to a minimum (Table S-5). This step, however, results in an increasing number of false positives which were reduced by the EIC filtering after integration of each feature using two fundamentally different algorithms. For both algorithms, various peak characteristics, such as peak height, peak width or retention time, were exported and used to create the filters (Table S-6). Also, relationships within the data obtained by one algorithm (e.g., difference between centroid and apex retention time) as well as between the data obtained by the two algorithms (e.g., relative changes in peak heights) were considered. Each filter consists of a minimum, a maximum and a variation argument. For the latter, samples were analyzed in triplicates[41] to regard the (relative) standard deviation, (R)SD. Peaks satisfying all filters were subjected to the blank correction, where a fold change (sample to blank) of five must be exceeded. A "componentization" step was conducted to group isotopologues, adducts and dimers (Table S-6).[33] The single steps of the entire workflow were merged in a Matlab script, which allows a semiautomated way of data processing including import and export data from/to vendor software via *.txt files. The processing strategy can be adapted to other instrument setups.

Derive Filter Settings from Training Data. For the assessment of water treatment processes, the reliability of the

data needs to be high. False positives/negatives hamper the evaluation by distorting the actual data and might lead to erroneous conclusions. To counter these problems, we developed an effective EIC filtering strategy to distinguish between true peaks and noise. A triplicate injection of a river water sample was used for the generation of training data. At this point, no blank correction was performed, as real chromatographic peaks also occur in the blank. Across the triplicates, 2893 features were recognized in at least one replicate. Each EIC was manually reviewed and categorized into the three groups *Peaks*, *No Peaks*, and *Exclude*. EICs exceeding the empiric intensity threshold of 200 cps and a signal-to-noise ratio (S/N) of about three were grouped into the category *Peaks*. EICs that did not exceed this S/N were assigned to *No Peaks*. Candidates with sufficient S/N but peak heights below the intensity threshold, as well as candidates which could not unambiguously assigned to *Peaks*/*No Peaks* were excluded from the data set in order to validate the filtering approach (true/false decision necessary). Likewise, implausible courses (e.g., strong variations, absence in some replicates) were assigned to *Exclude* and not considered any further. After manual review, the data arrays were split into *Peaks* and *No Peaks*. For each of the parameters, the distribution of the *Peaks* and *No Peaks* groups was taken into consideration. The absolute parameter as well as the fluctuation across the replicates was considered. For parameters subjected to a one-sided distribution (e.g., peak height), the 0.1% quantile was used as a lower threshold or the 99.9% quantile as an upper threshold. In case of two-sided distributions (e.g., peak width), the 0.05% and 99.95% quantiles were used. The derivation of the filter settings for the peak width is illustrated in Figure 1. The settings for the remaining filters were derived in the same way and are summarized in Table S-6. Some values are already determined by previous workflow steps and were therefore not derived from the training data.

Assessment of Water Treatment Processes. Since most of the peaks in nontarget screening remain unknown, the process performance needs to be based on signal heights rather than on concentrations. Changes in signal heights are theoretically mirrored by changes in concentrations, whereas matrix effects in electrospray ionization (in particular ion suppression) may affect this relation. Whether or to what extent these effects might impair the assessment depends on the degree of matrix change throughout the treatment process. In case of major changes of the sample matrix, the results should be interpreted with caution.[43] Recovery rates of spiked internal standards can serve as a rough estimate for the occurrence/extent of matrix effects, but accurate and precise correction methods—as is usual in target analysis—are future need.

To track features through the treatment process, the fold change (fc) between the effluent and influent sample was considered. For reasons of simplicity, the calculated fcs were classified to five distinct categories (Table 1). A signal decrease of more than 80% (fc <0.2) was defined as complete elimination. The interval for the respective counterpart (e.g., *elimination–formation*) was defined using its multiplicative inverse and thus led to fair classification irrespective of the direction being considered. The formation of a compound was consequently assumed for fc >5. The fc interval for the category *consistency* was derived from the stability of the internal standards by considering the maximum value of the 5-fold standard deviation.
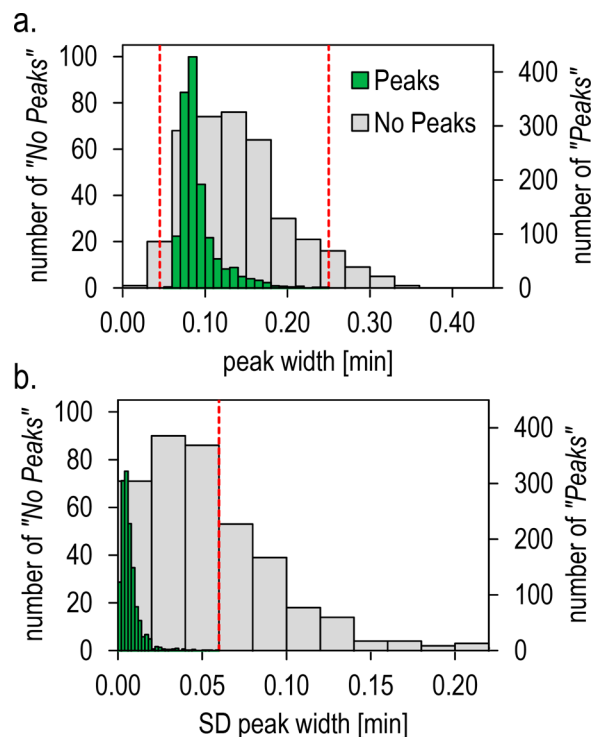

Figure 1. Distribution of (a) the peak widths (full width at half-maximum, fwhm) and (b) the standard deviations (SDs) of the peak widths across a triplicate for all *Peaks* and *No Peaks*. The bin widths were determined by the "auto" algorithm (histogram, Matlab). The derived filter settings (minimum = 0.045 min, maximum = 0.250 min, SD = 0.060 min) are depicted by the dashed lines.

Table 1. Peak Categorization Based on Defined Fold Changes in Peak Heights between Process Effluent and Influent

category	fc[a] interval
(E) elimination	$0.00 \leq fc < 0.20$
(D) decrease	$0.20 \leq fc < 0.50$
(C) consistency	$0.50 \leq fc \leq 2.00$
(I) increase	$2.00 < fc \leq 5.00$
(F) formation	$5.00 < fc \leq \infty$

[a]Fold change fc = peak height$_{Effluent}$/peak height$_{Influent}$

It should be noted that in addition to micropollutants, the signals of all detectable compounds (e.g., of natural origin) are considered for the assessment. Moreover, one component might generate multiple signals during the mass spectrometric detection. Some of those signals (e.g., isotopologues, adducts) can be grouped and assigned to one compound,[33] whereas others (e.g., in-source fragmentation) cannot easily be corrected. Redundancies in data interpretation are thus difficult to circumvent and hamper the numeric process assessment.

Repeatability of Classification. For evaluation of the repeatability of the signal classification, the ozonation process during drinking water treatment was considered. Both the process influent (river water) and effluent (ozonation effluent) were analyzed in nine technical replicates ($n = 9$). The presented data processing strategy requires the use of triplicates ($k = 3$). A combinatorial approach was conducted for the pairwise comparison of all possible combinations of triplicates out of n technical replicates. For estimating the repeatability of the complete workflow—including the fc categorization—two

aspects were taken into consideration: (a) The repeatability of technical replicates within one sample, meaning influent and effluent were composed of the same sample but different replicates. Considering this purely theoretical "process", each feature should be assigned to the category *consistency* (assuming the technical replicates to be identical). The pairwise comparison of triplicates out of n technical replicates of the same sample (Figure 2a) leads to a total number of
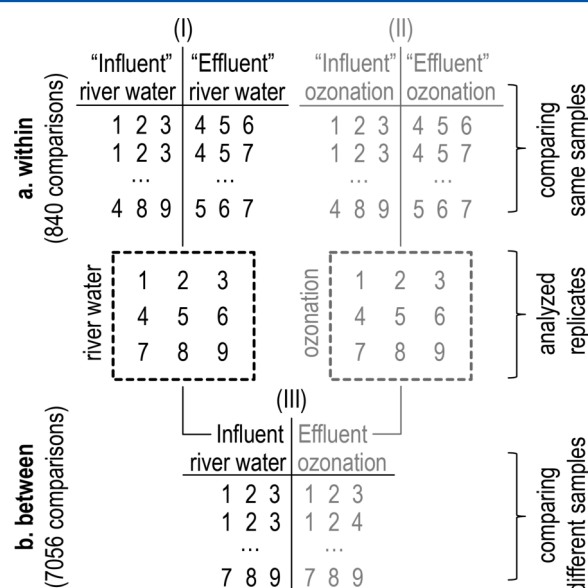

Figure 2. Pairwise comparison of triplicates ($k = 3$) out of $n = 9$ technical replicates (a) within same samples (theoretical process I and II) and (b) between different samples (real process of ozonation III).

$\frac{1}{2}\binom{n}{k}\binom{n-k}{k}$ "processes" to be considered (here 840). This evaluation was performed with the river water (I) and the effluent after ozonation (II) and served as a plausibility check. (b) The repeatability of technical replicates between two samples, meaning influent and effluent were composed of different samples (III). Considering this real process (ozonation), it was expected that changes caused by treatment were reflected by the number of features assigned to the five different categories (Table 1). The pairwise comparison of triplicates out of n technical replicates of the process influent and effluent

(Figure 2b) results in $\binom{n}{k}^2$ processes to be considered (here 7056).

RESULTS AND DISCUSSION

The first part describes the validation of the EIC filtering approach. The second part deals with data processing for the assessment of treatment processes. Specifically, we address the threshold problematic and present a recursive approach to overcome related problems. Finally, we evaluate the repeatability of the strategy using a combinatorial approach to demonstrate the general applicability.

Validation of EIC Filtering. The performance of the EIC filtering approach was assessed for its recall (i.e., true positive rate) and precision. The data obtained from peak finding and alignment was considered as a ground truth since the aim was to validate the EIC filtering step of the entire workflow. Four samples taken from the serial processes of the drinking water treatment served as a test data set. As with the training data, each feature was manually reviewed and categorized in *Peaks*, *No Peak*, or *Exclude*. After running through the nontargeted workflow including the derived EIC filtering criteria (except for the blank correction), the produced results (i.e., feature retained/rejected) were compared with the manual classification (i.e., *Peaks/No Peaks*) for determining common validation metrics. The summary of the validation for both the training and test data is listed in Table 2.

The good performance of the EIC filtering is underlined by values for precision and recall above 98% and 96%, respectively. The low numbers of false negative (fn) findings were mostly caused by integration problems of double peaks with insufficient chromatographic separation. Note that fn in this case refers to true peaks recognized by peak finding and alignment but rejected during the filtering approach. The false positive rate was always below 5%. Similar studies on wastewater reported higher proportions of fn and fp, which may be explained by stronger matrix interferences.[37,38]

In a second step, the performance of the EIC filtering was assessed based on 411 target compounds. A 0.1 $\mu g/L$ multicomponent standard (ultrapure water) was injected in triplicates (9.5 pg on column). The complete workflow was applied to the data set and targets were treated as unknown signals. The data set comprising 5435 features (recognized in at least one replicate) was searched against the target list (exact

Table 2. Validation of the EIC Filtering, Number of Features (without Blank Correction)

	training data	test data			
parameter	river water	river water	effluent O$_3$	effluent MLF	effluent ACF
		ground truth (manual decision)			
peaks	1376	1672	1502	1405	1317
no peaks	889	1225	1373	1452	1570
excluded	628	436	458	476	446
		validation			
true positive (tp)	1366	1648	1493	1397	1294
true negative (tn)	851	1172	1317	1404	1517
false positive (fp)	38	53	56	48	53
false negative (fn)	10	24	9	8	23
precision[a]	0.993	0.986	0.994	0.994	0.983
recall[b]	0.973	0.969	0.964	0.967	0.961
FPR[c]	0.043	0.043	0.041	0.033	0.034

[a]Precision = tp/(tp + fp). [b]Recall = tp/(tp + fn). [c]False positive rate FPR = fp/(fp + tn).

mass and expected retention time). Out of 411 compounds, 408 were successfully detected (at least in one replicate) in the peaks table after alignment. The three missing components all showed insufficient chromatographic separation with other compounds from the mix (Figure S-1). This underlines the importance of adequate separation methods which are difficult to optimize for such multimethods.[37] After applying the filters to the data, still 406 out of 408 possible components were retained. One was lost due to low signal height below the threshold, the second due to integration problems of a double peak. The recovery of more than 99.5% (406/408) emphasizes that the possibility of losing real peaks is deemed to be low. Considering the complete workflow, eight more standards were (correctly) removed due to the blank correction (fold change of sample to blank <5).

Threshold Problematic. In comparative sample evaluation, there are also cases where one peak (either influent or effluent) falls below the intensity threshold that need to be processed to determine the fold changes. In the following, this aspect is discussed for signals with decreasing peak heights over a treatment process. Generally, two different viewpoints exist. Peak heights below the intensity threshold can either be assumed to be zero (optimistic view) or be replaced by the intensity threshold itself (pessimistic view). Regardless of which view is considered to be the better option, inaccurate estimates will occur. Two features selected from the test data set emphasize possible misinterpretations that would occur if the optimistic (Figure S-2a) or the pessimistic view (Figure S-2b) was applied. Regarding the classification system, the optimistic view likely leads to overestimation of the groups *elimination/formation*, whereas the pessimistic one will generally favor the group *consistency*. The uncertainty problematic of the peak classification is depicted in Figure 3.
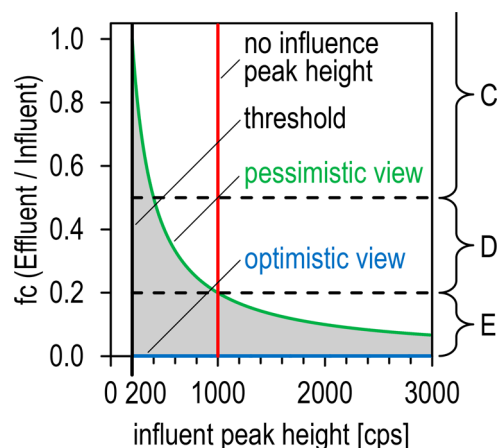

Figure 3. Uncertainty of the peak categorization based on the influent peak height (*abscissa*) and the intensity threshold used (here 200 cps). The predefined intervals for the groups *elimination* (E), *decrease* (D), and *consistency* (C) are indicated by the dashed lines. The gray area highlights the uncertainty region.

The threshold of 200 cps and the categories E, D, and C are used for the discussion. Assuming an influent peak height of 300 cps and an effluent peak height below the threshold, the signal could theoretically be assigned to all three categories: C (e.g., 200/300, pessimistic), D (e.g., 120/300), or E (e.g., 0/300, optimistic). The correct choice remains unknown and the arbitrary assignment leads to blurred boundaries which might

result in misinterpretations of the factual circumstances. After exceeding the influent peak height of 1000 cps, the classification is no longer affected by the threshold since the uncertainty region completely falls within category E. Thus, in this case only features with peak heights above 1000 cps (in influent or effluent) can be unambiguously assigned to the defined groups. However, more than 75% of all features from the training and test data sets (after EIC filtering) fall within the region between the threshold and the "no influence peak height" of 1000 cps (Figure S-3). Instead of discarding large parts of the data for the assessment of a process, we developed a recursive approach for more reliable assignment of features within the uncertainty region.

Recursive Approach for Classification. Two algorithms were applied for the peak integration for two different reasons: First, multiple valuable filtering criteria could be derived to effectively reduce false positives. Second, the so-called *Summation* algorithm forces EIC integration irrespective of absence/presence of a chromatographic peak. We found this aspect to be helpful for a reliable sample comparison based on fcs. For one feature, the integration by the *Summation* algorithm is shown for the process influent and effluent (Figure 4). The algorithm first detects the apex within the defined
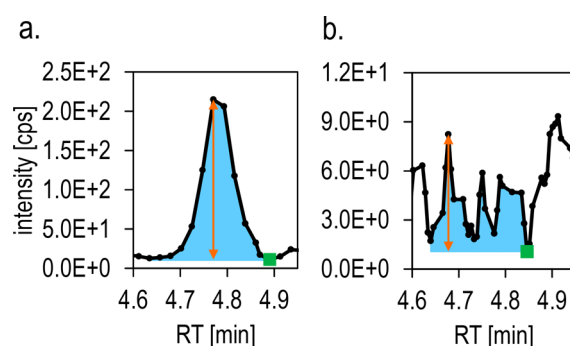

Figure 4. EICs of a feature in a. the influent (river water) and b. the effluent (ACF effluent) sample of a treatment process integrated by the *Summation* algorithm. The data point with minimum abundance within the summation window is displayed as a square and the peak height is indicated by an arrow.

summation window. The data point with minimum abundance within the summation window is used to estimate the horizontal baseline. Finally, the end points of the summation range are adjusted to local minima (if any). The actual height is the intensity difference between the apex and the baseline estimate. The heights reported by this algorithm were recursively retrieved for fc calculations involving signals that were initially rejected (e.g., below the intensity threshold). This recursive step, however, was only applied to features which were confirmed by the EIC filtering in one sample (influent or effluent) but rejected in the other. In such cases, the data obtained by the *Summation* algorithm was retrieved for the rejected triplicate, which did not satisfy the filter settings. If the SD of the summation peak heights across this triplicate was higher than half the intensity threshold, the fluctuation criterion (RSD < 0.5, Table S-6) needed to be fulfilled. For SDs below half the intensity threshold or RSDs below 0.5, the summation peak height was used for the fc calculations. Otherwise, the feature was rejected due to the high fluctuation which would lead to inconclusive classification.

The recursive approach allows a more reliable classification of signal changes without skewing the results by arbitrary assumptions concerning signals below the threshold. The influent peak height (204 cps) shown in Figure 4 is only slightly above the threshold and thus falls within the uncertainty region for the classification. Considering the determined height (here to be seen as noise) in the effluent sample (7 cps), the signal can be assigned to the group E (7/204 < 0.20). For chromatograms with high noise fluctuation, the height determined by the *Summation* algorithm can be quite high and may in some instances lead to wrong classification. Especially if the noise level between influent and effluent sample changes drastically, such influences are conceivable.

For the blank correction, the same recursive strategy involving the *Summation* algorithm was applied. The blanks were not subjected to the EIC filtering approach but rather compared to signals which met all derived criteria. Features were rejected if neither the influent nor the effluent satisfied the fc criteria relative to the blank (fc >5). The blank correction was less prone to errors using this approach.

Evaluating the Repeatability. The aim was to evaluate the reliability of the presented data processing strategy. The similarity of the results (i.e., classification of features) was assessed on the basis of multiple technical replicates (see Methods section). As a proof of concept, we used the ozonation process during drinking water treatment. As a first step, the system stability across the batch was evaluated by reviewing the spiked IS (Figure S-4). Problems such as loss of sensitivity over time (here >14 h runtime) would impair the feature classification and thus result in poor repeatability. The good system stability was indicated by peak height RSDs always below 10%. Furthermore, the impact of matrix effects on the signal classification can be estimated by comparing IS peak heights between the different matrices. The pairwise comparison of triplicates from the influent and effluent sample resulted in IS fcs ranging between 0.88 and 1.17 (Figure S-5), which clearly fell within the interval of the category C. Thus, all 13 IS would always be assigned to C (without blank correction). We therefore assumed that sensitivity changes or matrix effects—in this case—had negligible influence on the feature classification.

In the next step, the combinatorial approach was conducted to compare all possible combinations of technical replicates within (840) and between (7056) the two samples. For each single comparison of triplicates (e.g., replicates 1, 2, and 3 compared with nos. 4, 5, and 6), the number of features assigned to the five categories were reported. However, for evaluation of the repeatability, each single feature had to be considered as the sums of all features assigned to certain categories could be repeatable while the individual compositions of the sums were not. Thus, in addition to the mere number of features assigned to a category, we also reviewed how many features were always (i.e., in all possible comparisons) assigned to this category. The two different perspectives considered during data processing are schematically depicted in Figure S-6.

After data evaluation, two different types of features were distinguished: Features unambiguously assigned to one category and features ambiguously assigned to multiple categories. The manual inspection of the latter revealed that problems were mainly caused by three different reasons:

(1) Signals were in vicinity of the intensity threshold or the blank threshold. With some signals slightly above/below the threshold, these features were sometimes rejected during EIC filtering while other combinations of replicates passed the filtering approach. To overcome these problems, the processing (during this evaluation) was repeated for signals classified to multiple categories using less stringent criteria in a second step. The minimum values for the peak height and area were reduced by 50%, the factor for blank correction reduced to 3. All other filtering criteria (e.g., RSDs) were not modified.

(2) The calculated fc was in vicinity of the upper or lower limit of the fc interval used for classification. Across all combinations, some were slightly above or below this limit resulting in ambiguous assignment of the feature into two adjacent categories (e.g., *elimination* and *decrease*). The classification algorithm was therefore adapted to handle such cases. Features for which the number of all possible comparisons (840 or 7056) was reached by counting two adjacent categories were unambiguously assigned to the category comprising the larger number of comparisons.

(3) Real peaks showed implausible signal courses across the batch (strong fluctuations, trends). These features reflect real differences within replicates of the same sample (unexplained circumstances) and are not caused by data processing. A reasonable correction of such profiles is not possible.

Representative examples of peak profiles illustrating these problem cases are summarized in Figure S-7.

The results of the data evaluation with respect to the repeatability are summarized in Figure 5.
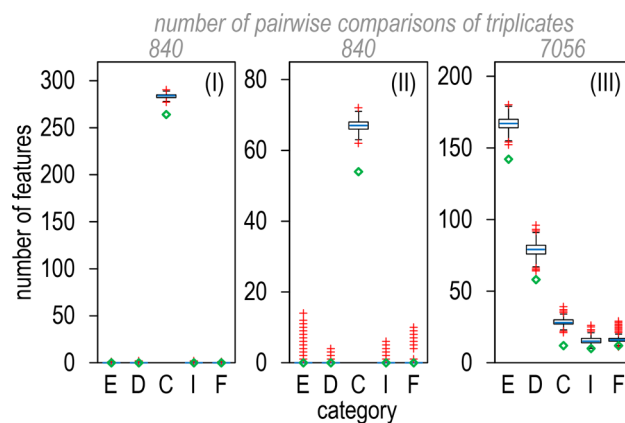

Figure 5. Boxplots of the number of features assigned to the categories elimination (E), decrease (D), consistency (C), increase (I), or formation (F) across all possible comparisons within one sample (river water, I and ozonation effluent, II) and between two samples (real process of ozonation, III). Outliers shown as pluses, diamonds indicate the number of features which were unambiguously assigned to the respective category. For more details, see Figure S-8.

The lacking ground truth (i.e., true feature classification not known) in nontarget analysis requires alternative approaches for plausibility checks of generated results. Here, the within sample comparisons (Figure 5, I and II) were conducted to illustrate that differences between technical replicates (theoretical process) hardly exist. As was to be expected, features were almost exclusively assigned to C and only few outliers were classified to the remaining categories (75th percentile was zero for E, D, I, and F). The higher uncertainties observed for the effluent samples (II) may be explained by lower feature intensities. Regarding category C, the low fluctuation of the assigned number of features throughout all comparisons is indicated by the length of the boxplots. While this suggests a

high repeatability, there is a certain offset between features unambiguously assigned to *C* and the distribution of all individual comparisons. The majority of these differences was caused by implausible signal courses (discussed above) or interfering peaks (either insufficient chromatographic or mass spectrometric separation). Notably, most of the ambiguously assigned features were detected in retention time regions where the total ion chromatograms (of influent and effluent) suggest high probability of interferences (Figure S-9). Optimization of the chromatographic separation might improve this situation.

Considering the ozonation (Figure 5, III), strong changes between influent and effluent sample became apparent. Again, the narrow distribution of the number of features assigned to the respective categories points to good repeatability. The differences between those distributions and the number of features unambiguously assigned to one category can be explained by the same reasons as stated above. The treatment efficiency of the ozonation is indicated by the large proportion of features assigned to E and D. Considering the median values across all conducted comparisons, 55% of all signals were assigned to E, 26% to D, 9% to C, 5% to I, and 5% to F. However, it was expected that more transformation products (features assigned to F) were formed during the ozonation. Even though the entire window of detectable components is strongly extended using nontarget screening for the assessment of processes, there are still many compounds which cannot be captured by the acquisition method used. Each single step, that is, separation, ionization, detection, as well as data processing, requires certain criteria which narrow down the diversity of the substances being considered.[44] Previous research on ozonation with nontarget screening reported similar relative numbers subjected to elimination, whereas more formation has been detected for the ozonation of wastewater.[30] A more comprehensive interpretation of the results was not within the scope of this study.

In addition to existing validation strategies for untargeted methods (e.g., refs 37 and 45), this combinatorial approach might help to better evaluate the reliability of the final outcome.

CONCLUSION

In this study, a data processing strategy for the assessment of water treatment processes was developed. The validation of the workflow suggests that false positive/negative findings have negligible influence on the results.

A concept for process assessment based on signal changes across the treatment was introduced. The use of isotope-labeled internal standards is strongly recommended to estimate the influence of matrix effects, which have to be seen with particular caution. The recursive data processing allowed a more reliable estimate of fold changes. We further recommend applying the same recursive integration for a less error prone blank correction. The main limitations of the signal classification represent (i) redundancies in data interpretation as one component might generate multiple features (e.g., in-source fragmentation) and (ii) implausible signal courses across the measured batch. Further efforts are necessary to keep the influence of the mentioned problems to a minimum.

The evaluation of the entire processing strategy by using a combinatorial approach indicates sufficient repeatability. We, therefore, conclude that nontarget screening using LC-HRMS in combination with appropriate data processing of sample triplicates provides a more holistic picture of changes throughout processes and can thus reduce existing knowledge

gaps. Monitoring of water treatment processes or optimization of different operating conditions are, inter alia, fields of application, which could benefit from this more comprehensive view.

The processing strategies and—more importantly—the combinatorial validation concept may be easily transferred to other research fields. In a figurative sense, the term *process* should be seen as abstract expression for a two-sample comparison.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.7b03037.

List of 411 target compounds, smiles code for target compounds with missing CAS numbers, LC-MS acquisition method, list of isotope-labeled standards, parameters for data processing, optimized parameters for EIC filtering, problem cases occurring during peak alignment, examples for possible misinterpretations caused by fc calculation, cumulative intensity distribution of the training data and the test data, stability of isotope-labeled standards, boxplots of the fold changes for the spiked isotope-labeled standards, data processing strategy for assessing the repeatability, peak profiles to illustrate the classification problems, boxplots created from the repeatability considerations, and m/z-RT-distribution of ambiguously assigned features (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: bader.t@lw-online.de. Phone: + 49 7345 9638 2865. Fax: + 49 7345 9638 2290.

ORCID Ⓞ
Tobias Bader: 0000-0002-7274-5015

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Schwarzenbach, R. P.; Escher, B. I.; Fenner, K.; Hofstetter, T. B.; Johnson, C. A.; von Gunten, U.; Wehrli, B. *Science* **2006**, *313*, 1072−1077.
(2) Richardson, S. D.; Ternes, T. A. *Anal. Chem.* **2014**, *86*, 2813−2848.
(3) Howard, P. H.; Muir, D. C. G. *Environ. Sci. Technol.* **2010**, *44*, 2277−2285.
(4) Reemtsma, T.; Weiss, S.; Mueller, J.; Petrovic, M.; Gonzalez, S.; Barcelo, D.; Ventura, F.; Knepper, T. P. *Environ. Sci. Technol.* **2006**, *40*, 5451−5458.
(5) Seitz, W.; Winzenbacher, R. *Environ. Monit. Assess.* **2017**, *189*, 244.
(6) Loos, R.; Carvalho, R.; António, D. C.; Comero, S.; Locoro, G.; Tavazzi, S.; Paracchini, B.; Ghiani, M.; Lettieri, T.; Blaha, L.; Jarosova, B.; Voorspoels, S.; Servaes, K.; Haglund, P.; Fick, J.; Lindberg, R. H.; Schwesig, D.; Gawlik, B. M. *Water Res.* **2013**, *47*, 6475−6487.
(7) Castronovo, S.; Wick, A.; Scheurer, M.; Nödler, K.; Schulz, M.; Ternes, T. A. *Water Res.* **2017**, *110*, 342−353.

(8) Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P. *Water Res.* **2015**, *87*, 145−154.

(9) Schlüsener, M. P.; Kunkel, U.; Ternes, T. A. *Environ. Sci. Technol.* **2015**, *49*, 14282−14291.

(10) Fleischer, S.; Weiss, S. C.; Lucke, T.; Seitz, W.; Schulz, W.; Weber, W. H. *Ozone: Sci. Eng.* **2015**, *37*, 441−449.

(11) Zahn, D.; Frömel, T.; Knepper, T. P. *Water Res.* **2016**, *101*, 292−299.

(12) Herrmann, M.; Menz, J.; Olsson, O.; Kümmerer, K. *Water Res.* **2015**, *85*, 11−21.

(13) Segura, P. A.; MacLeod, S. L.; Lemoine, P.; Sauvé, S.; Gagnon, C. *Chemosphere* **2011**, *84*, 1085−1094.

(14) Yang, Y.; Ok, Y. S.; Kim, K.-H.; Kwon, E. E.; Tsang, Y. F. *Sci. Total Environ.* **2017**, *596*, 303−320.

(15) Prasse, C.; Stalter, D.; Schulte-Oehlmann, U.; Oehlmann, J.; Ternes, T. A. *Water Res.* **2015**, *87*, 237−270.

(16) Knopp, G.; Prasse, C.; Ternes, T. A.; Cornel, P. *Water Res.* **2016**, *100*, 580−592.

(17) Margot, J.; Kienle, C.; Magnet, A.; Weil, M.; Rossi, L.; de Alencastro, L. F.; Abegglen, C.; Thonney, D.; Chèvre, N.; Schärer, M.; Barry, D. A. *Sci. Total Environ.* **2013**, *461−462*, 480−498.

(18) Jekel, M.; Dott, W.; Bergmann, A.; Dünnbier, U.; Gnirss, R.; Haist-Gulde, B.; Hamscher, G.; Letzel, M.; Licha, T.; Lyko, S.; Miehe, U.; Sacher, F.; Scheurer, M.; Schmidt, C. K.; Reemtsma, T.; Ruhl, A. S. *Chemosphere* **2015**, *125*, 155−167.

(19) Moschet, C.; Wittmer, I.; Simovic, J.; Junghans, M.; Piazzoli, A.; Singer, H.; Stamm, C.; Leu, C.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 5423−5432.

(20) Petrie, B.; Barden, R.; Kasprzyk-Hordern, B. *Water Res.* **2015**, *72*, 3−27.

(21) Krauss, M.; Singer, H.; Hollender, J. *Anal. Bioanal. Chem.* **2010**, *397*, 943−951.

(22) Hernández, F.; Sancho, J. V.; Ibáñez, M.; Abad, E.; Portolés, T.; Mattioli, L. *Anal. Bioanal. Chem.* **2012**, *403*, 1251−1264.

(23) Moschet, C.; Piazzoli, A.; Singer, H.; Hollender, J. *Anal. Chem.* **2013**, *85*, 10312−10320.

(24) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M.; Onghena, M.; Kunkel, U.; Schulz, W.; et al. *Anal. Bioanal. Chem.* **2015**, *407*, 6237−6255.

(25) Müller, A.; Schulz, W.; Ruck, W. K. L.; Weber, W. H. *Chemosphere* **2011**, *85*, 1211−1219.

(26) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. *Environ. Pollut.* **2014**, *184*, 25−32.

(27) Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. *Anal. Chem.* **2015**, *87*, 12121−12129.

(28) Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. *Environ. Sci. Technol.* **2015**, *49*, 12333−12341.

(29) Muz, M.; Ost, N.; Kühne, R.; Schüürmann, G.; Brack, W.; Krauss, M. *Chemosphere* **2017**, *166*, 300−310.

(30) Merel, S.; Lege, S.; Yanez Heras, J. E.; Zwiener, C. *Environ. Sci. Technol.* **2017**, *51*, 410−417.

(31) Vergeynst, L.; Van Langenhove, H.; Demeestere, K. *Anal. Chem.* **2015**, *87*, 2170−2177.

(32) Samanipour, S.; Reid, M. J.; Thomas, K. V. *Anal. Chem.* **2017**, *89*, 5585−5591.

(33) Loos, M. Dissertation, ETH Zurich, 2015. DOI: 10.3929/ethz-a-010645125.

(34) Schymanski, E. L.; Singer, H. P.; Longrée, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripollés Vidal, C.; Hollender, J. *Environ. Sci. Technol.* **2014**, *48*, 1811−1818.

(35) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. *Environ. Sci. Technol.* **2017**, *51*, 11505−11512.

(36) Merel, S.; Anumol, T.; Park, M.; Snyder, S. A. J. *J. Hazard. Mater.* **2015**, *282*, 75−85.

(37) Nürenberg, G.; Schulz, M.; Kunkel, U.; Ternes, T. A. *J. Chromatogr. A* **2015**, *1426*, 77−90.

(38) Parry, E.; Young, T. M. *Water Res.* **2016**, *104*, 72−81.

(39) Bader, T.; Schulz, W.; Lucke, T.; Seitz, W.; Winzenbacher, R. In *Assessing Transformation Products of Chemicals by Non-Target and Suspect Screening—Strategies and Workflows Vol. 2*; American Chemical Society, 2016; pp 49−70.

(40) Heuett, N. V. Dissertation, Florida International University, 2015; http://digitalcommons.fiu.edu/etd/2194.

(41) Bader, T.; Schulz, W.; Kümmerer, K.; Winzenbacher, R. *Anal. Chim. Acta* **2016**, *935*, 173−186.

(42) Fawcett, T. *Pattern Recognit. Lett.* **2006**, *27*, 861−874.

(43) Kruve, A.; Herodes, K.; Leito, I. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1159−1168.

(44) Reemtsma, T.; Berger, U.; Arp, H. P. H.; Gallard, H.; Knepper, T. P.; Neumann, M.; Quintana, J. B.; Voogt, P. *Environ. Sci. Technol.* **2016**, *50*, 10308−10315.

(45) Naz, S.; Vallejo, M.; García, A.; Barbas, C. *J. Chromatogr. A* **2014**, *1353*, 99−105.

Supporting Information for:

LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes

Tobias Bader[1,2*], Wolfgang Schulz[1], Klaus Kümmerer[2], Rudi Winzenbacher[1]

[1]*Laboratory for Operation Control and Research, Zweckverband Landeswasserversorgung, Am Spitzigen Berg 1, 89129 Langenau, Germany*
[2]*Sustainable Chemistry and Material Resources, Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg, Scharnhorststraße 1/C13, 21335 Lüneburg, Germany*

*Corresponding author
Email: Bader.T@lw-online.de
Phone: +49 7345 9638 2279
Fax: +49 7345 9638 2290

TABLE OF CONTENTS (clickable links)

Table S-1. List of 411 target compounds

#	Name	Formula	CAS Number	#	Name	Formula	CAS Number
1	1-(2-Chlorophenyl)-3-phenylurea	$C_{13}H_{11}ClN_2O$	2989-99-3	41	Aminocarb	$C_{11}H_{16}N_2O_2$	2032-59-9
2	1-(3,4-Dichlorophenyl)-3-methylurea	$C_8H_8Cl_2N_2O$	3567-62-2	42	Amisulprid	$C_{17}H_{27}N_3O_4S$	71675-85-9
3	1-(3,4-Dichlorophenyl)urea	$C_7H_6Cl_2N_2O$	2327-02-8	43	Amisulpride N-Oxide	$C_{17}H_{27}N_3O_5S$	71676-01-2
4	1,1-Dimethyl-3-phenylurea	$C_9H_{12}N_2O$	101-42-8	44	Aspartame	$C_{14}H_{18}N_2O_5$	22839-47-0
5	1,2,3-Benzotriazole	$C_6H_5N_3$	95-14-7	45	Asulam	$C_8H_{10}N_2O_4S$	3337-71-1
6	1-Hydroxyisoquinoline	C_9H_7NO	491-30-5	46	Atenolol	$C_{14}H_{22}N_2O_3$	29122-68-7
7	1-Methylisoquinoline	$C_{10}H_9N$	1721-93-3	47	Atraton	$C_9H_{17}N_5O$	1610-17-9
8	2-(Methylthio)benzothiazole	$C_8H_7NS_2$	615-22-5	48	Atrazine	$C_8H_{14}ClN_5$	1912-24-9
9	2,4-Dimethylquinoline	$C_{11}H_{11}N$	1198-37-4	49	Atrazine-2-hydroxy	$C_8H_{15}N_5O$	2163-68-0
10	2,5-Dichloroaniline	$C_6H_5Cl_2N$	95-82-9	50	Azamethiphos	$C_9H_{10}ClN_2O_5PS$	35575-96-3
11	2,6-Dichlorbenzamid	$C_7H_5Cl_2NO$	2008-58-4	51	Azinphos-ethyl	$C_{12}H_{16}N_3O_3PS_2$	2642-71-9
12	2-Aminobenzothiazole	$C_7H_6N_2S$	136-95-8	52	Azoxystrobin	$C_{22}H_{17}N_3O_5$	131860-33-8
13	2-Hydroxybenzothiazole	C_7H_5NOS	934-34-9	53	Benalaxyl-M	$C_{20}H_{23}NO_3$	98243-83-5
14	2-Methoxyphenyl isocyanate	$C_8H_7NO_2$	700-87-8	54	Benazolin	$C_9H_6ClNO_3S$	3813-05-6
15	2-Methyl-8-quinolinol	$C_{10}H_9NO$	826-81-3	55	Bensulfuron-methyl	$C_{16}H_{18}N_4O_7S$	83055-99-6
16	2-Methylphenyl isocyanate	C_8H_7NO	614-68-6	56	Bentazon	$C_{10}H_{12}N_2O_3S$	25057-89-0
17	3-(2-Chloro-6-methylphenyl)-1,1-dimethylurea	$C_{10}H_{13}ClN_2O$	15441-90-4	57	Benzocaine	$C_9H_{11}NO_2$	94-09-7
18	3-(Trifluoromethyl)aniline	$C_7H_6F_3N$	98-16-8	58	Benzothiazole-6-carboxylic acid	$C_8H_5NO_2S$	3622-35-3
19	3-Hydroxycarbofuran	$C_{12}H_{15}NO_4$	16655-82-6	59	Betaxolol	$C_{18}H_{29}NO_3$	63659-18-7
20	4-Acetamidoantipyrine	$C_{13}H_{15}N_3O_2$	83-15-8	60	Bezafibrate	$C_{19}H_{20}ClNO_4$	41859-67-0
21	4-Chloro-o-toluidine	C_7H_8ClN	95-69-2	61	Bisoprolol	$C_{18}H_{31}NO_4$	66722-44-9
22	4'-Hydroxydiclofenac	$C_{14}H_{11}Cl_2NO_3$	64118-84-9	62	Bisoprolol_M1	$C_{15}H_{23}NO_5$	109791-19-7
23	4-Isopropylaniline	$C_9H_{13}N$	99-88-7	63	Bisoprolol_M3	$C_{13}H_{19}NO_4$	72570-70-8
24	5,6-Dimethyl-1H-benzotriazole	$C_8H_9N_3$	4184-79-6	64	Bisoprolol_M4	$C_{15}H_{25}NO_4$	109791-18-6
25	5'-Chloro-2'-methylacetanilide	$C_9H_{10}ClNO$	5900-55-0	65	Bitertanol	$C_{20}H_{23}N_3O_2$	55179-31-2
26	5-Methyl-1H-benzotriazole	$C_7H_7N_3$	136-85-6	66	Boscalid	$C_{18}H_{12}Cl_2N_2O$	188425-85-6
27	8-Hydroxyquinoline	C_9H_7NO	148-24-3	67	Bromacil	$C_9H_{13}BrN_2O_2$	314-40-9
28	Acebutolol	$C_{18}H_{28}N_2O_4$	37517-30-9	68	Bromophos-ethyl	$C_{10}H_{12}BrCl_2O_3PS$	4824-78-6
29	Acephate	$C_4H_{10}NO_3PS$	30560-19-1	69	Bromuconazole	$C_{13}H_{12}BrCl_2N_3O$	116255-48-2
30	Acetaminophen	$C_8H_9NO_2$	103-90-2	70	Bupirimate	$C_{13}H_{24}N_4O_3S$	41483-43-6
31	Acetamiprid	$C_{10}H_{11}ClN_4$	135410-20-7	71	Caffeine	$C_8H_{10}N_4O_2$	58-08-2
32	Acetylsulfamethoxazole	$C_{12}H_{13}N_3O_4S$	21312-10-7	72	Candesartan	$C_{24}H_{20}N_6O_3$	139481-59-7
33	Aclonifen	$C_{12}H_9ClN_2O_3$	74070-46-5	73	Carbamazepine	$C_{15}H_{12}N_2O$	298-46-4
34	Acridine	$C_{13}H_9N$	260-94-6	74	Carbamazepine 10,11-epoxide	$C_{15}H_{12}N_2O_2$	36507-30-9
35	Alachlor	$C_{14}H_{20}ClNO_2$	15972-60-8	75	Carbanilide	$C_{13}H_{12}N_2O$	102-07-8
36	Aldicarb-sulfoxide	$C_7H_{14}N_2O_3S$	1646-87-3	76	Carbaryl	$C_{12}H_{11}NO_2$	63-25-2
37	Alprenolol	$C_{15}H_{23}NO_2$	13655-52-2	77	Carbendazim	$C_9H_9N_3O_2$	10605-21-7
38	Amantadine	$C_{10}H_{17}N$	768-94-5	78	Carbetamide	$C_{12}H_{16}N_2O_3$	16118-49-3
39	Ametryn	$C_9H_{17}N_5S$	834-12-8	79	Carbofuran	$C_{12}H_{15}NO_3$	1563-66-2
40	Amidosulfuron	$C_9H_{15}N_5O_7S_2$	120923-37-7	80	Chloramben	$C_7H_5Cl_2NO_2$	133-90-4

#	Name	Formula	CAS Number
81	Chlorbromuron	$C_9H_{10}BrClN_2O_2$	13360-45-7
82	Chlordimeform	$C_{10}H_{13}ClN_2$	6164-98-3
83	Chlorfenvinphos	$C_{12}H_{14}Cl_3O_4P$	470-90-6
84	Chloridazon	$C_{10}H_8ClN_3O$	1698-60-8
85	Chloridazon-methyl-desphenyl	$C_5H_6ClN_3O$	17254-80-7
86	Chloroxuron	$C_{15}H_{15}ClN_2O_2$	1982-47-4
87	Chlorpyrifos	$C_9H_{11}Cl_3NO_3PS$	2921-88-2
88	Chlorthalonil R611965	$C_8H_4Cl_3NO_3$	142733-37-7
89	Chlortoluron	$C_{10}H_{13}ClN_2O$	15545-48-9
90	Chlortoluron benzoic acid (CGA 15140)	$C_{10}H_{11}ClN_2O_3$	59587-01-8
91	Clarithromycin	$C_{38}H_{69}NO_{13}$	81103-11-9
92	Clenbuterol	$C_{12}H_{18}Cl_2N_2O$	37148-27-9
93	Clindamycin	$C_{18}H_{33}ClN_2O_5S$	18323-44-9
94	Clodinafop-propargyl	$C_{17}H_{13}ClFNO_4$	105512-06-9
95	Clomazone	$C_{12}H_{14}ClNO_2$	81777-89-1
96	Clothianidin	$C_6H_8ClN_5O_2S$	210880-92-5
97	Codeine	$C_{18}H_{21}NO_3$	76-57-3
98	Crotamiton	$C_{13}H_{17}NO$	483-63-6
99	Cyanazine	$C_9H_{13}ClN_6$	21725-46-2
100	Cybutryne	$C_{11}H_{19}N_5S$	28159-98-0
101	Cyproconazole	$C_{15}H_{18}ClN_3O$	94361-06-5
102	Cyprodinil	$C_{14}H_{15}N_3$	121552-61-2
103	Dapsone	$C_{12}H_{12}N_2O_2S$	80-08-0
104	Deisopropylatrazine	$C_5H_8ClN_5$	1007-28-9
105	Desethylatrazine	$C_6H_{10}ClN_5$	6190-65-4
106	Desethylterbutylazine	$C_7H_{12}ClN_5$	30125-63-4
107	Diatrizoate	$C_{11}H_9I_3N_2O_4$	737-31-5
108	Diazepam	$C_{16}H_{13}ClN_2O$	439-14-5
109	Diazinon	$C_{12}H_{21}N_2O_3PS$	333-41-5
110	Dichlofenthion	$C_{10}H_{13}Cl_2O_3PS$	97-17-6
111	Dichlorvos	$C_4H_7Cl_2O_4P$	62-73-7
112	Diclofenac	$C_{14}H_{11}Cl_2NO_2$	15307-86-5
113	Dicrotophos	$C_8H_{16}NO_5P$	141-66-2
114	Didesmethylisoproturon	$C_{10}H_{14}N_2O$	56046-17-4
115	Diethofencarb	$C_{14}H_{21}NO_4$	87130-20-9
116	Difenoconazole	$C_{19}H_{17}Cl_2N_3O_3$	119446-68-3
117	Difenoxuron	$C_{16}H_{18}N_2O_3$	14214-32-5
118	Diflubenzuron	$C_{14}H_9ClF_2N_2O_2$	35367-38-5
119	Diflufenican	$C_{19}H_{11}F_5N_2O_2$	83164-33-4
120	Dihydrocodeine	$C_{18}H_{23}NO_3$	125-28-0
121	Dimefuron	$C_{15}H_{19}ClN_4O_3$	34205-21-5
122	Dimethachlor	$C_{13}H_{18}ClNO_2$	50563-36-5

#	Name	Formula	CAS Number
123	Dimethachlor CGA 102935	$C_{12}H_{13}NO_5$	[1)]
124	Dimethachlor CGA 354742	$C_{13}H_{19}NO_5S$	[1)]
125	Dimethachlor CGA 369873	$C_{10}H_{13}NO_4S$	[1)]
126	Dimethachlor CGA 373464	$C_{12}H_{15}NO_6S$	1196157-87-5
127	Dimethachlor CGA 50266	$C_{13}H_{17}NO_4$	1086384-49-7
128	Dimethachlor SYN 528702	$C_{15}H_{21}NO_5S$	1228182-52-2
129	Dimethachlor SYN 530561	$C_{13}H_{17}NO_5$	1138220-18-4
130	Dimethenamid	$C_{12}H_{18}ClNO_2S$	87674-68-8
131	Dimethenamid M23	$C_{12}H_{17}NO_4S$	380412-59-9
132	Dimethenamid M27	$C_{12}H_{19}NO_5S_2$	205939-58-8
133	Dimethenamid-P M31	$C_{14}H_{21}NO_5S_2$	[1)]
134	Dimethoate	$C_5H_{12}NO_3PS_2$	60-51-5
135	Dimethomorph	$C_{21}H_{22}ClNO_4$	113210-97-2
136	Dimoxystrobin	$C_{19}H_{22}N_2O_3$	149961-52-4
137	Dimoxystrobin 505M08 (BF 505-7)	$C_{19}H_{20}N_2O_5$	[1)]
138	Dimoxystrobin 505M09 (BF 505-8)	$C_{19}H_{20}N_2O_5$	[1)]
139	Diniconazole	$C_{15}H_{17}Cl_2N_3O$	83657-24-3
140	Diphenylamine	$C_{12}H_{11}N$	122-39-4
141	Disulfoton-sulfone	$C_8H_{19}O_4PS_3$	2497-06-5
142	Disulfoton-sulfoxide	$C_8H_{19}O_3PS_3$	2497-07-6
143	Ditalimfos	$C_{12}H_{14}NO_4PS$	5131-24-8
144	Diuron	$C_9H_{10}Cl_2N_2O$	330-54-1
145	Doxycycline	$C_{22}H_{24}N_2O_8$	564-25-0
146	Epoxiconazole	$C_{17}H_{13}ClFN_3O$	135319-73-2
147	Eprosartan	$C_{23}H_{24}N_2O_4S$	133040-01-4
148	Erythromycin	$C_{37}H_{67}NO_{13}$	114-07-8
149	Estrone	$C_{18}H_{22}O_2$	53-16-7
150	Ethenzamide	$C_9H_{11}NO_2$	938-73-8
151	Ethidimuron	$C_7H_{12}N_4O_3S_2$	30043-49-3
152	Ethion	$C_9H_{22}O_4P_2S_4$	563-12-2
153	Ethirimol	$C_{11}H_{19}N_3O$	23947-60-6
154	Ethofumesate	$C_{13}H_{18}O_5S$	26225-79-6
155	Ethoprophos	$C_8H_{19}O_2PS_2$	13194-48-4
156	Etofibrate	$C_{18}H_{18}ClNO_5$	31637-97-5
157	Fenamiphos sulfone	$C_{13}H_{22}NO_5PS$	31972-44-8
158	Fenarimol	$C_{17}H_{12}Cl_2N_2O$	60168-88-9
159	Fenazaquin	$C_{20}H_{22}N_2O$	120928-09-8
160	Fenbuconazole	$C_{19}H_{17}ClN_4$	114369-43-6
161	Fenhexamid	$C_{14}H_{17}Cl_2NO_2$	126833-17-8
162	Fenitrothion	$C_9H_{12}NO_5PS$	122-14-5
163	Fenobucarb	$C_{12}H_{17}NO_2$	3766-81-2
164	Fenofibrate	$C_{20}H_{21}ClO_4$	49562-28-9

#	Name	Formula	CAS Number
165	Fenoxaprop	$C_{16}H_{12}ClNO_5$	95617-09-7
166	Fenoxycarb	$C_{17}H_{19}NO_4$	72490-01-8
167	Fenpropidin	$C_{19}H_{31}N$	67306-00-7
168	Fenpropimorph	$C_{20}H_{33}NO$	67564-91-4
169	Fenpyroximate	$C_{24}H_{27}N_3O_4$	111812-58-9
170	Fenthion	$C_{10}H_{15}O_3PS_2$	55-38-9
171	Fipronil	$C_{12}H_4Cl_2F_6N_4OS$	120068-37-3
172	Flamprop	$C_{16}H_{13}ClFNO_3$	58667-63-3
173	Flazasulfuron	$C_{13}H_{12}F_3N_5O_5S$	104040-78-0
174	Flecainide	$C_{17}H_{20}F_6N_2O_3$	54143-55-4
175	Florasulam	$C_{12}H_8F_3N_5O_3S$	145701-23-1
176	Fluazifop	$C_{15}H_{12}F_3NO_4$	69335-91-7
177	Fluazinam	$C_{13}H_4Cl_2F_6N_4O_4$	79622-59-6
178	Flufenacet	$C_{14}H_{13}F_4N_3O_2S$	142459-58-3
179	Flufenacet ESA	$C_{11}H_{14}FNO_4S$	201668-32-8
180	Flufenoxuron	$C_{21}H_{11}ClF_6N_2O_3$	101463-69-8
181	Fluometuron	$C_{10}H_{11}F_3N_2O$	2164-17-2
182	Fluopicolide	$C_{14}H_8Cl_3F_3N_2O$	239110-15-7
183	Fluquinconazole	$C_{16}H_8Cl_2FN_5O$	136426-54-5
184	Fluroxypyr	$C_7H_5Cl_2FN_2O_3$	69377-81-7
185	Flurtamone	$C_{18}H_{14}F_3NO_2$	96525-23-4
186	Flusilazole	$C_{16}H_{15}F_2N_3Si$	85509-19-9
187	Foramsulfuron	$C_{17}H_{20}N_6O_7S$	173159-57-4
188	Gabapentin	$C_9H_{17}NO_2$	60142-96-3
189	Gabapentin-Lactam	$C_9H_{15}NO$	64744-50-9
190	Haloxyfop	$C_{15}H_{11}ClF_3NO_4$	69806-34-4
191	Heptenophos	$C_9H_{12}ClO_4P$	23560-59-0
192	Hexa(methoxymethyl)melamine	$C_{15}H_{30}N_6O_6$	3089-11-0
193	Hexaconazole	$C_{14}H_{17}Cl_2N_3O$	79983-71-4
194	Hexazinone	$C_{12}H_{20}N_4O_2$	51235-04-2
195	Hexythiazox	$C_{17}H_{21}ClN_2O_2S$	78587-05-0
196	Ifosfamide	$C_7H_{15}Cl_2N_2O_2P$	3778-73-2
197	Imazalil	$C_{14}H_{14}Cl_2N_2O$	35554-44-0
198	Imazapyr	$C_{13}H_{15}N_3O_3$	81334-34-1
199	Imazaquin	$C_{17}H_{17}N_3O_3$	81335-37-7
200	Imidaclopride	$C_9H_{10}ClN_5O_2$	138261-41-3
201	Indomethacin	$C_{19}H_{16}ClNO_4$	53-86-1
202	Indoxacarb	$C_{22}H_{17}ClF_3N_3O_7$	173584-44-6
203	Iodofenphos	$C_8H_8Cl_2IO_3PS$	18181-70-9
204	Iodosulfuron-methyl	$C_{14}H_{14}IN_5O_6S$	185119-76-0
205	Iohexol	$C_{19}H_{26}I_3N_3O_9$	66108-95-0
206	Iomeprol	$C_{17}H_{22}I_3N_3O_8$	78649-41-9

#	Name	Formula	CAS Number
207	Iopromide	$C_{18}H_{24}I_3N_3O_8$	73334-07-3
208	Iprodione	$C_{13}H_{13}Cl_2N_3O_3$	36734-19-7
209	Iprovalicarb	$C_{18}H_{28}N_2O_3$	140923-17-7
210	Irbesartan	$C_{25}H_{28}N_6O$	138402-11-6
211	Irbesartan_446	$C_{25}H_{30}N_6O_2$	1)
212	Isoprocarb	$C_{11}H_{15}NO_2$	2631-40-5
213	Isoproturon	$C_{12}H_{18}N_2O$	34123-59-6
214	Ketoprofen	$C_{16}H_{14}O_3$	22071-15-4
215	Ketotifen	$C_{19}H_{19}NOS$	34580-14-8
216	Kresoxim (BF 490-1)	$C_{17}H_{17}NO_4$	137169-29-0
217	Kresoxim-methyl	$C_{18}H_{19}NO_4$	143390-89-0
218	Lamotrigine N2-Oxide	$C_9H_7Cl_2N_5O$	136565-76-9
219	Linuron	$C_9H_{10}Cl_2N_2O_2$	330-55-2
220	Losartan	$C_{22}H_{23}ClN_6O$	114798-26-4
221	Malaoxon	$C_{10}H_{19}O_7PS$	1634-78-2
222	Malathion	$C_{10}H_{19}O_6PS_2$	121-75-5
223	Mecarbam	$C_{10}H_{20}NO_5PS_2$	2595-54-2
224	Mefenpyr-diethyl	$C_{16}H_{18}Cl_2N_2O_4$	135590-91-9
225	Mepanipyrim	$C_{14}H_{13}N_3$	110235-47-7
226	Mephosfolan	$C_8H_{16}NO_3PS_2$	950-10-7
227	Mepronil	$C_{17}H_{19}NO_2$	55814-41-0
228	Mesosulfuron-methyl	$C_{17}H_{21}N_5O_9S_2$	208465-21-8
229	Mestranol	$C_{21}H_{26}O_2$	72-33-3
230	Metalaxyl	$C_{15}H_{21}NO_4$	57837-19-1
231	Metalaxyl-M CGA 108906	$C_{14}H_{17}NO_6$	104390-56-9
232	Metalaxyl-M CGA 62826	$C_{14}H_{19}NO_4$	75596-99-5
233	Metamitron	$C_{10}H_{10}N_4O$	41394-05-2
234	Metamitron-desamino	$C_{10}H_9N_3O$	36993-94-9
235	Metazachlor	$C_{14}H_{16}ClN_3O$	67129-08-2
236	Metazachlor BH 479-11	$C_{15}H_{19}N_3O_2S$	1)
237	Metazachlor BH 479-12	$C_{14}H_{13}N_3O_5$	1)
238	Metazachlor BH 479-4	$C_{14}H_{15}N_3O_3$	1231244-60-2
239	Metazachlor BH 479-8	$C_{14}H_{17}N_3O_4S$	172960-62-2
240	Metazachlor BH 479-9	$C_{16}H_{19}N_3O_4S$	1)
241	Metconazole	$C_{17}H_{22}ClN_3O$	125116-23-6
242	Methabenzthiazuron	$C_{10}H_{11}N_3O_S$	18691-97-9
243	Methidathion	$C_6H_{11}N_2O_4PS_3$	950-37-8
244	Methiocarb	$C_{11}H_{15}NO_2S$	2032-65-7
245	Methiocarb sulfoxide	$C_{11}H_{15}NO_3S$	2635-10-1
246	Methiocarb-sulfone	$C_{11}H_{15}NO_4S$	2179-25-1
247	Methoxyfenozide	$C_{22}H_{28}N_2O_3$	161050-58-4
248	Methylphenidate	$C_{14}H_{19}NO_2$	113-45-1

#	Name	Formula	CAS Number	#	Name	Formula	CAS Number
249	Metobromuron	$C_9H_{11}BrN_2O_2$	3060-89-7	291	Oxazepam	$C_{15}H_{11}ClN_2O_2$	604-75-1
250	Metolachlor	$C_{15}H_{22}ClNO_2$	51218-45-2	292	Oxydemeton-methyl	$C_6H_{15}O_5PS_2$	301-12-2
251	Metolachlor CGA 354743	$C_{15}H_{23}NO_5S$	171118-09-5	293	Oxytetracycline	$C_{22}H_{24}N_2O_9$	2058-46-0
252	Metolachlor CGA 357704	$C_{14}H_{17}NO_5$	1217465-10-5	294	Paraoxon	$C_{10}H_{14}NO_6P$	311-45-5
253	Metolachlor CGA 37735	$C_{11}H_{15}NO_2$	97055-05-5	295	Paraoxon-methyl	$C_8H_{10}NO_6P$	950-35-6
254	Metolachlor CGA 50267	$C_{12}H_{17}NO_2$	82508-03-0	296	Parathion	$C_{10}H_{14}NO_5PS$	56-38-2
255	Metolachlor CGA 50720	$C_{11}H_{13}NO_3$	152019-74-4	297	Parathion-methyl	$C_8H_{10}NO_5PS$	298-00-0
256	Metolachlor CGA 51202	$C_{15}H_{21}NO_4$	152019-73-3	298	Penconazole	$C_{13}H_{15}Cl_2N_3$	66246-88-6
257	Metolachlor NOA 413173	$C_{14}H_{19}NO_6S$	[1)]	299	Pencycuron	$C_{19}H_{21}ClN_2O$	66063-05-6
258	Metolcarb	$C_9H_{11}NO_2$	1129-41-5	300	Pendimethalin	$C_{13}H_{19}N_3O_4$	40487-42-1
259	Metoprolol	$C_{15}H_{25}NO_3$	51384-51-1	301	Pentoxifylline	$C_{13}H_{18}N_4O_3$	6493-05-6
260	Metoprolol acid	$C_{14}H_{21}NO_4$	56392-14-4	302	Pethoxamid	$C_{16}H_{22}ClNO_2$	106700-29-2
261	Metosulam	$C_{14}H_{13}Cl_2N_5O_4S$	139528-85-1	303	Pethoxamid sulphonic acid (MET-42)	$C_{16}H_{23}NO_5S$	1329805-71-1
262	Metoxuron	$C_{10}H_{13}ClN_2O_2$	19937-59-8	304	Phenacetin	$C_{10}H_{13}NO_2$	62-44-2
263	Metribuzin	$C_8H_{14}N_4OS$	21087-64-9	305	Phenanthridinon	$C_{13}H_9NO$	1015-89-0
264	Metronidazole	$C_6H_9N_3O_3$	443-48-1	306	Phenazone	$C_{11}H_{12}N_2O$	60-80-0
265	Metsulfuron-methyl	$C_{14}H_{15}N_5O_6S$	74223-64-6	307	Phenmedipham	$C_{16}H_{16}N_2O_4$	13684-63-4
266	Metyrapone	$C_{14}H_{14}N_2O$	54-36-4	308	Phenthoate	$C_{12}H_{17}O_4PS_2$	2597-03-7
267	Monalide	$C_{13}H_{18}ClNO$	7287-36-7	309	Phenylalanine	$C_9H_{11}NO_2$	150-30-1
268	Monocrotophos	$C_7H_{14}NO_5P$	6923-22-4	310	Phenylethylmalonamide	$C_{11}H_{14}N_2O_2$	7206-76-0
269	Monodemethylisoproturon	$C_{11}H_{16}N_2O$	34123-57-4	311	Phosalone	$C_{12}H_{15}ClNO_4PS_2$	2310-17-0
270	Monolinuron	$C_9H_{11}ClN_2O_2$	1746-81-2	312	Phosphamidon	$C_{10}H_{19}ClNO_5P$	13171-21-6
271	Monuron	$C_9H_{11}ClN_2O$	150-68-5	313	Phoxim	$C_{12}H_{15}N_2O_3PS$	14816-18-3
272	Myclobutanil	$C_{15}H_{17}ClN_4$	88671-89-0	314	Picloram	$C_6H_3Cl_3N_2O_2$	1918-02-1
273	N,N-Diethyltoluamide	$C_{12}H_{17}NO$	134-62-3	315	Picolinafen	$C_{19}H_{12}F_4N_2O_2$	137641-05-5
274	N,N-Dimethylaniline	$C_8H_{11}N$	121-69-7	316	Picoxystrobin	$C_{18}H_{16}F_3NO_4$	117428-22-5
275	Naphazoline	$C_{14}H_{14}N_2$	835-31-4	317	Pilocarpine	$C_{11}H_{16}N_2O_2$	92-13-7
276	Napropamide	$C_{17}H_{21}NO_2$	15299-99-7	318	Pindolol	$C_{14}H_{20}N_2O_2$	13523-86-9
277	Naproxen	$C_{14}H_{14}O_3$	22204-53-1	319	Piperophos	$C_{14}H_{28}NO_3PS_2$	24151-93-7
278	N-Ethylaniline	$C_8H_{11}N$	103-69-5	320	Pirimicarb	$C_{11}H_{18}N_4O_2$	23103-98-2
279	N-Formyl-4-aminoantipyrin	$C_{12}H_{13}N_3O_2$	1672-58-8	321	Pirimiphos-ethyl	$C_{13}H_{24}N_3O_3PS$	23505-41-1
280	Nicosulfuron	$C_{15}H_{18}N_6O_6S$	111991-09-4	322	Pirimiphos-methyl	$C_{11}H_{20}N_3O_3PS$	29232-93-7
281	N-Methyl-2-pyrrolidone	C_5H_9NO	872-50-4	323	Pregabalin	$C_8H_{17}NO_2$	148553-50-8
282	N-Methylbenzene-sulfonamide	$C_7H_9NO_2S$	5183-78-8	324	Prilocaine	$C_{13}H_{20}N_2O$	721-50-6
283	Norethisterone acetate	$C_{22}H_{28}O_3$	51-98-9	325	Primidone	$C_{12}H_{14}N_2O_2$	125-33-7
284	Norflurazon	$C_{12}H_9ClF_3N_3O$	27314-13-2	326	Primisulfuron-methyl	$C_{15}H_{12}F_4N_4O_7S$	86209-51-0
285	Nuarimol	$C_{17}H_{12}ClFN_2O$	63284-71-9	327	Prochloraz	$C_{15}H_{16}Cl_3N_3O_2$	67747-09-5
286	Olmesartan	$C_{24}H_{26}N_6O_3$	144689-24-7	328	Procymidone	$C_{13}H_{11}Cl_2NO_2$	32809-16-8
287	Omethoate	$C_5H_{12}NO_4PS$	1113-02-6	329	Profenofos	$C_{11}H_{15}BrClO_3PS$	41198-08-7
288	Oseltamivir	$C_{16}H_{28}N_2O_4$	196618-13-0	330	Promecarb	$C_{12}H_{17}NO_2$	2631-37-0
289	Oxadixyl	$C_{14}H_{18}N_2O_4$	77732-09-3	331	Prometon	$C_{10}H_{19}N_5O$	1610-18-0
290	Oxasulfuron	$C_{17}H_{18}N_4O_6S$	144651-06-9	332	Propamocarb	$C_9H_{20}N_2O_2$	24579-73-5

#	Name	Formula	CAS Number	#	Name	Formula	CAS Number
333	Propaquizafop	$C_{22}H_{22}ClN_3O_5$	111479-05-1	375	TCMTB	$C_9H_6N_2S_3$	21564-17-0
334	Propazine	$C_9H_{16}ClN_5$	139-40-2	376	Tebuconazol	$C_{16}H_{22}ClN_3O$	107534-96-3
335	Propazine-2-hydroxy	$C_9H_{17}N_5O$	7374-53-0	377	Tebufenpyrad	$C_{18}H_{24}ClN_3O$	119168-77-3
336	Propetamphos	$C_{10}H_{20}NO_4PS$	31218-83-4	378	Tebutam	$C_{15}H_{23}NO$	35256-85-0
337	Propiconazole	$C_{15}H_{17}Cl_2N_3O_2$	60207-90-1	379	Telmisartan	$C_{33}H_{30}N_4O_2$	144701-48-4
338	Propoxur	$C_{11}H_{15}NO_3$	114-26-1	380	Terbuthylazin-desethyl-2-hydroxy	$C_7H_{13}N_5O$	66753-06-8
339	Propranolol	$C_{16}H_{21}NO_2$	13013-17-7	381	Terbuthylazine	$C_9H_{16}ClN_5$	5915-41-3
340	Propyphenazone	$C_{14}H_{18}N_2O$	479-92-5	382	Terbutryn	$C_{10}H_{19}N_5S$	886-50-0
341	Prosulfocarb	$C_{14}H_{21}NOS$	52888-80-9	383	Terbutylazin 1 SYN 545666	$C_8H_{14}N_4O_2$	[1]
342	Prosulfuron	$C_{15}H_{16}F_3N_5O_4S$	94125-34-5	384	Terbutylazin 2 CGA 324007	$C_7H_{12}N_4O_2$	309923-18-0
343	Pyridaben	$C_{19}H_{25}ClN_2OS$	96489-71-3	385	Tetraconazole	$C_{13}H_{11}Cl_2F_4N_3O$	112281-77-3
344	Pyridate	$C_{19}H_{23}ClN_2O_2S$	55512-33-9	386	Tetramethylurea	$C_5H_{12}N_2O$	632-22-4
345	Pyrifenox	$C_{14}H_{12}Cl_2N_2O$	88283-41-4	387	Thiacloprid	$C_{10}H_9ClN_4S$	111988-49-9
346	Pyriproxyfen	$C_{20}H_{19}NO_3$	95737-68-1	388	Thiamethoxam	$C_8H_{10}ClN_5O_3S$	153719-23-4
347	Pyroquilon	$C_{11}H_{11}NO$	57369-32-1	389	Thifensulfuron-methyl	$C_{12}H_{13}N_5O_6S_2$	79277-27-3
348	Quinmerac	$C_{11}H_8ClNO_2$	90717-03-6	390	Tolfenamic acid	$C_{14}H_{12}ClNO_2$	13710-19-5
349	Quinmerac BH 518-2	$C_{11}H_6ClNO_4$	90717-07-0	391	Topramezone	$C_{16}H_{17}N_3O_5S$	210631-68-8
350	Quinoline	C_9H_7N	91-22-5	392	Tramadol	$C_{16}H_{25}NO_2$	27203-92-5
351	Quinoxyfen	$C_{15}H_8Cl_2FNO$	124495-18-7	393	Tramadol N-Oxide	$C_{16}H_{25}NO_3$	147441-56-3
352	Rimsulfuron	$C_{14}H_{17}N_5O_7S_2$	122931-48-0	394	trans-10,11-Dihydro-10,11-dihydroxy Carbamazepine	$C_{15}H_{14}N_2O_3$	58955-93-4
353	Ritalinic acid	$C_{13}H_{17}NO_2$	19395-41-6	395	Triadimenol	$C_{14}H_{18}ClN_3O_2$	89482-17-7
354	Ronidazol	$C_6H_8N_4O_4$	7681-76-7	396	Triallate	$C_{10}H_{16}Cl_3NOS$	2303-17-5
355	Roxithromycin	$C_{41}H_{76}N_2O_{15}$	80214-83-1	397	Triasulfuron	$C_{14}H_{16}ClN_5O_5S$	82097-50-5
356	Schradan	$C_8H_{24}N_4O_3P_2$	152-16-9	398	Triazophos	$C_{12}H_{16}N_3O_3PS$	24017-47-8
357	Sebuthylazine	$C_9H_{16}ClN_5$	7286-69-3	399	Triethyl phosphate	$C_6H_{15}O_4P$	78-40-0
358	Sebuthylazine-desethyl	$C_7H_{12}ClN_5$	37019-18-4	400	Trifloxistrobin CGA 321113	$C_{19}H_{17}F_3N_2O_4$	252913-85-2
359	Simazine	$C_7H_{12}ClN_5$	122-34-9	401	Trifloxistrobin NOA 413161	$C_{19}H_{15}F_3N_2O_6$	[1]
360	Simazine-2-hydroxy	$C_7H_{13}N_5O$	2599-11-3	402	Trifloxystrobin	$C_{20}H_{19}F_3N_2O_4$	141517-21-7
361	Simeton	$C_8H_{15}N_5O_1$	673-04-1	403	Triflusulfuron-methyl	$C_{17}H_{19}F_3N_6O_6S$	126535-15-7
362	Simetryn	$C_8H_{15}N_5S$	1014-70-6	404	Trimethoprim	$C_{14}H_{18}N_4O_3$	738-70-5
363	Sitagliptin	$C_{16}H_{15}F_6N_5O$	486460-32-6	405	Tritosulfuron	$C_{13}H_9F_6N_5O_4S$	142469-14-5
364	S-Metolachlor Metabolite CGA 368208	$C_{11}H_{15}NO_4S$	1173021-76-5	406	Tritosulfuron BH 635-4	$C_{10}H_{10}F_3N_5O_4S$	[1]
365	Sotalol	$C_{12}H_{20}N_2O_3S$	3930-20-9	407	Tritosulfuron BH 635-5	$C_5H_5F_3N_4O$	5311-05-7
366	Spiroxamine	$C_{18}H_{35}NO_2$	118134-30-8	408	Valsartan	$C_{24}H_{29}N_5O_3$	137862-53-4
367	Sulfadiazine	$C_{10}H_{10}N_4O_2S$	68-35-9	409	Valsartan acid	$C_{14}H_{10}N_4O_2$	164265-78-5
368	Sulfadimidine	$C_{12}H_{14}N_4O_2S$	57-68-1	410	Warfarin	$C_{19}H_{16}O_4$	81-81-2
369	Sulfamerazine	$C_{11}H_{12}N_4O_2S$	127-79-7	411	Xanthone	$C_{13}H_8O_2$	90-47-1
370	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	723-46-6				
371	Sulfathiazole	$C_9H_9N_3O_2S_2$	72-14-0				
372	Sulpirid	$C_{15}H_{23}N_3O_4S$	15676-16-1				
373	Sulpiride N-Oxide	$C_{15}H_{23}N_3O_5S$	[1]				
374	Swep	$C_8H_7Cl_2NO_2$	1918-18-9				

[1]: For substances with missing CAS number, the smiles code is given in **Table S-2**

Table S-2. Smiles code for target compounds with missing CAS numbers

#[1]	Name	Formula	Smiles code
123	Dimethachlor CGA 102935	$C_{12}H_{13}NO_5$	Cc1cccc(c1N(CC(=O)O)C(=O)C(=O)O)C
124	Dimethachlor CGA 354742	$C_{13}H_{19}NO_5S$	Cc1cccc(c1N(CCOC)C(=O)CS(=O)(=O)O)C
125	Dimethachlor CGA 369873	$C_{10}H_{13}NO_4S$	Cc1cccc(c1NC(=O)CS(=O)(=O)O)C
133	Dimethenamid-P M31	$C_{14}H_{21}NO_5S_2$	Cc1csc(c1N(C(C)COC)C(=O)CS(=O)CC(=O)O)C
137	Dimoxystrobin 505M08 (BF 505-7)	$C_{19}H_{20}N_2O_5$	Cc1ccc(c(c1)OCc2ccccc2/C(=N\OC)/C(=O)NC)C(=O)O
138	Dimoxystrobin 505M09 (BF 505-8)	$C_{19}H_{20}N_2O_5$	CNC(=O)C(=N\OC)\c1ccccc1COc1cc(ccc1C)C(O)=O
211	Irbesartan_446	$C_{25}H_{30}N_6O_2$	CCCCC(=O)NC1(CCCC1)C(=O)NCc2ccc(cc2)c3ccccc3c4[nH]nnn4
236	Metazachlor BH 479-11	$C_{15}H_{19}N_3O_2S$	Cc1cccc(C)c1N(Cn1cccn1)C(=O)CS(C)=O
237	Metazachlor BH 479-12	$C_{14}H_{13}N_3O_5$	Cc1cccc(C(O)=O)c1N(Cn1cccn1)C(=O)C(O)=O
240	Metazachlor BH 479-9	$C_{16}H_{19}N_3O_4S$	Cc1cccc(C)c1N(Cn1cccn1)C(=O)CS(=O)CC(O)=O
257	Metolachlor NOA 413173	$C_{14}H_{19}NO_6S$	CCc1cccc(c1N(C(C)C(=O)O)C(=O)CS(=O)(=O)O)C
373	Sulpiride N-Oxide	$C_{15}H_{23}N_3O_5S$	CC[N+]1(CCCC1CNC(=O)C2=C(C=CC(=C2)S(=O)(=O)N)OC)[O-]
383	Terbutylazin 1 SYN 545666	$C_8H_{14}N_4O_2$	CC(C)(C)Nc1[nH]c(=O)n(c(=O)n1)C
401	Trifloxystrobin NOA 413161	$C_{19}H_{15}F_3N_2O_6$	CO\N=C(\C(O)=O)c1ccccc1CO\N=C(/C(O)=O)c1cccc(c1)C(F)(F)F
406	Tritosulfuron BH 635-4	$C_{10}H_{10}F_3N_5O_4S$	NC(=O)NC(=N)NC(=O)NS(=O)(=O)c1ccccc1C(F)(F)F

[1]: Numbers taken from **Table S-1**

Table S-3. LC-MS acquisition method

Name	
LC System	LC20 series Shimadzu
LC column	Zorbax Eclipse Plus C18 (2.1 x 150 mm, 3.5 µm) Agilent
Mobile phase A	Ultrapure water + 0.1 % v/v formic acid
Mobile phase B	Acetonitrile + 0.1 % v/v formic acid
Gradient (%B)	0 min (2%), 1 min (2%), 2 min (20%), 16.5 min (100%), 27 min (100%), 27.1 min (2%), 37 min (2%)
Flow rate	0.3 mL/min
Temperature	40 °C
Sample injection volume	95 µL
Internal standard injection volume	5 µL
MS System	TripleTOF^{TM} 5600 (Sciex)
Ion source	DuoSpray Ion Source
Mass range	m/z 100 - 1200
Survey scan accumulation time	250 ms
Cycle time	ca. 1.1 sec[1]
Ion Source Gas 1	35 psi
Ion Source Gas 2	45 psi
Curtain Gas	40 psi
Source Temperature	550 °C
IonSpray Voltage Floating	5500 V
Declustering Potential	60 V
Collision Energy	10 eV (for MS mode)

[1]: In addition to MS, also MS² experiments have been performed (data dependent acquisition as well as data independent acquisition - SWATH^{TM}). The results of the study, however, are only based on the LC-MS data (survey scan). MS² acquisition was conducted to receive realistic cycle times comparable to approaches used in reality (where MS² is of great interest)

Table S-4. List of isotope-labeled standards (IS)

Name	Formula	Retention time [min]	Concentration[1] [µg L^{-1}]
Benzotriazole D4	$C_6 H N_3 {}^2H_4$	5.5	10.0
Chloridazon D5	$C_{10} H_3 Cl N_3 O {}^2H_5$	6.4	0.20
Propazine D6	$C_9 H_{10} Cl N_5 {}^2H_6$	10.7	0.20
Diuron D6	$C_9 H_4 Cl_2 N_2 O {}^2H_6$	9.6	0.60
Lidocaine D10	$C_{14} H_{12} N_2 O {}^2H_{10}$	5.3	0.10
Sotalol D6	$C_{12} H_{14} N_2 O_3 S {}^2H_6$	4.5	0.40
Hydrochlorothiazide 13C, D2	$C_6 H_6 Cl N_3 O_4 S_2 {}^{13}C {}^2H_2$	5.1	40.0
Diazinon D10	$C_{12} H_{11} N_2 O_3 P S {}^2H_{10}$	13.7	0.10
Sulfadimethoxine D6	$C_{12} H_8 N_4 O_4 S {}^2H_6$	7.5	0.40
Azoxystrobin D4	$C_{22} H_{13} N_3 O_5 {}^2H_4$	11.5	0.10
Irbesartan D4	$C_{25} H_{24} N_6 O {}^2H_4$	8.7	0.40
Bicalutamide D4	$C_{18} H_{10} F_4 N_2 O_4 S {}^2H_4$	10.7	0.50
Darunavir D9	$C_{27} H_{28} N_3 O_7 S {}^2H_9$	10.2	20.0

[1]: Due to varying ionization efficiency, different concentrations were used to receive similar signal abundances

Table S-5. Parameters for data processing (*workflow LW_V2.0*)

Software package	Parameter	Value
	Peak finding	
	Minimum retention time	3.5 min
	Maximum retention time	19.0 min
	Subtraction offset	6 scans
	Subtraction multiple factor	1.3
	Noise threshold	25 cps
MarkerView™	Minimum spectral peak width	20 ppm
(1.2.1)	Minimum RT peak width	3 scans
	Assign charge states	Enabled
	Peak alignment	
	Retention time tolerance	0.15 min
	Mass tolerance	20 ppm
	Intensity threshold	100 cps
	Peak integration by MQ4 algorithm	
	Gaussian Smooth Width	1.0 points
	RT Half Window	9 sec
	Update Expected RT	No
	Report Largest Peak	No
	Min. Peak Width	6 points
	Min. Peak Height	100
MultiQuant™ 2.1.2	Noise Percentage	90.0%
+ Research Feature	Baseline Sub. Window	0.3 min
Software (Jan 2017)	Peak Splitting	2
	Peak integration by Summation (SUM) algorithm	
	Gaussian Smooth Width	1.0 points
	Summation Window	15 sec
	Noise% for Baseline	-1[1]
	Recentering	3.0
	Adjust Endpoints to Local Minima	Yes

Parameters for peak finding and peak alignment were optimized in a former study (data not shown).

[1]: Option which is exclusively available in the research version of MultiQuant™, the data point with the smallest intensity within the summation window is used to fit the horizontal baseline of the peak

Table S-6. Optimized parameters for EIC filtering (*workflow LW_V2.0*)

Software package	Parameter	Value			
		min	max	RSD[1]	SD[1]
	MQ4 algorithm				
	(1) Expected RT[2] [min]	**3.5**	**19**	-	-
	(2) Peak Area [cts]	**700**	**Inf**	0.50	-
	(3) Peak Height [cps]	**200**	**Inf**	0.45	-
	(4) Centroid RT[3] [min]	**3.5**	**19**	-	0.07
	(5) Peak Width (FWHM) [min]	0.045	0.250	-	0.06
	(6) Baseline Delta[4] [ ]	0	0.35	-	0.25
	(7) Apex RT[5] [min]	**3.5**	**19**	-	0.08
	(8) Δ (Expected RT, Centroid RT) [min]	**-0.15**	**0.15**	-	0.08
	(9) Δ (Expected RT, Apex RT) [min]	**-0.15**	**0.15**	-	0.08
	(10) Δ (Centroid RT, Apex RT) [min]	-0.04	0.04	-	0.05
	Summation (SUM) algorithm				
Matlab R2016b	(11) Expected RT[2] [min]	**3.5**	**19**	-	-
	(12) Peak Area [cts]	**700**	**Inf**	0.60	-
	(13) Peak Height [cps]	**200**	**Inf**	0.50	-
	(14) Centroid RT[3] [min]	**3.5**	**19**	-	0.07
	(15) Peak Width (FWHM) [min]	0.050	0.250	-	0.06
	(16) Apex RT[5] [min]	**3.5**	**19**	-	0.07
	(17) Δ(Expected RT, Centroid RT) [min]	**-0.15**	**0.15**	-	0.07
	(18) Δ(Expected RT, Apex RT) [min]	**-0.15**	**0.15**	-	0.07
	(19) Δ(Centroid RT, Apex RT) [min]	-0.04	0.04		0.06
	Between both algorithms				
	(20) Δ(Centroid RT MQ4, SUM) [min]	-0.06	0.06	-	0.05
	(21) Δ(Apex RT MQ4, SUM) [min]	-0.10	0.10	-	0.07
	(22) Δ(Peak Width MQ4, SUM) [min]	-0.09	0.09	-	0.07
	(23) Δ_{rel}(Peak Height MQ4, SUM)[6] [ ]	-0.30	0.55	-	0.40

[1]: Both the RSD and the SD are calculated across the triplicates; *note: RSD is dimensionless*

[2]: Expected RT: initially set retention time value for a feature prior to integration (taken from peak alignment)

[3]: Centroid RT: intensity weighted average retention time across a chromatographic peak (analogous to centroid mass)

[4]: Baseline Delta: height difference of the baseline (at start and end of the peak) to the actual peak height

[5]: Apex RT: retention time value reported for the most abundant data point across a chromatographic peak

[6]: Δ_{rel} = (Peak Height (MQ4) - Peak Height (SUM)) / mean(Peak Height (MQ4), Peak Height (SUM))

bold printed: parameters were defined and not derived from the training data set; either empiric (e.g. peak height) or already determined by other settings (e.g. +/- 0.15 min taken from peak alignment)

Note: due to inadequate noise estimation in the MultiQuant software, the signal-to-noise ratio was not considered as a useful filter.

Componentization was conducted within 0.15 min and 20 ppm tolerances for the following species: 1st and 2nd isotope, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, $[M+CH_3CN+H]^+$, $[2M+H]^+$, $[2M+Na]^+$, $[2M+K]^+$, $[2M+NH_4]^+$.
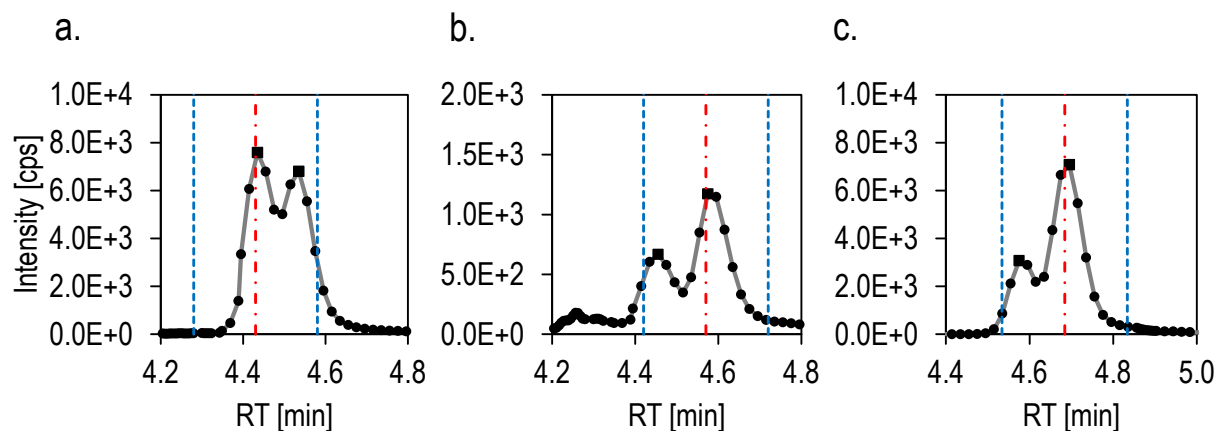
Figure S-1. Problem cases occurring during peak alignment. The smaller peak was not listed in the final peak list as the retention time tolerance during peak alignment (indicated by blue lines +/- of detected apex) includes the apex of the smaller peak. a. **Aminocarb** (2nd) and Pilocarpine (both $C_{11}H_{16}N_2O_2$), b. **N,N-Dimethylaniline** (1st) and N-Ethylaniline (both $C_8H_{11}N$), c. Simazine-2-hydroxy (2nd) and **Terbuthylazine-desethyl-2-hydroxy** (both $C_7H_{13}N_5O$), bold printed compounds were missed due to peak alignment.
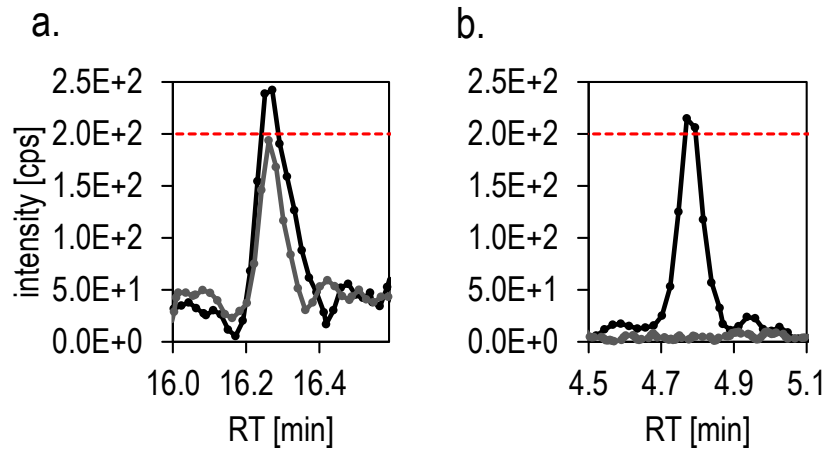
Figure S-2. Examples which emphasize possible misinterpretations that would occur if the fc calculations were conducted using a. the optimistic or b. the pessimistic view. Extracted ion chromatograms (EIC) of the influent (black) and effluent (grey) sample of a treatment process. The intensity threshold is indicated by the dashed lines.
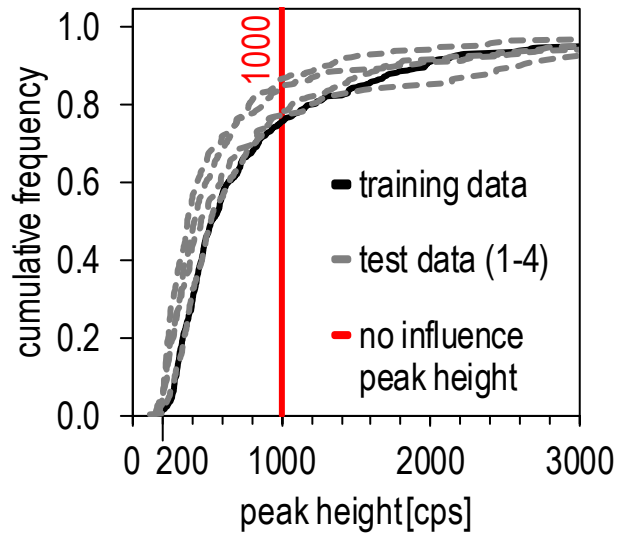
Figure S-3. Cumulative intensity distribution of the training data and the test data (sample 1-4). The "no influence peak height" of 1000 cps is depicted by the red line.
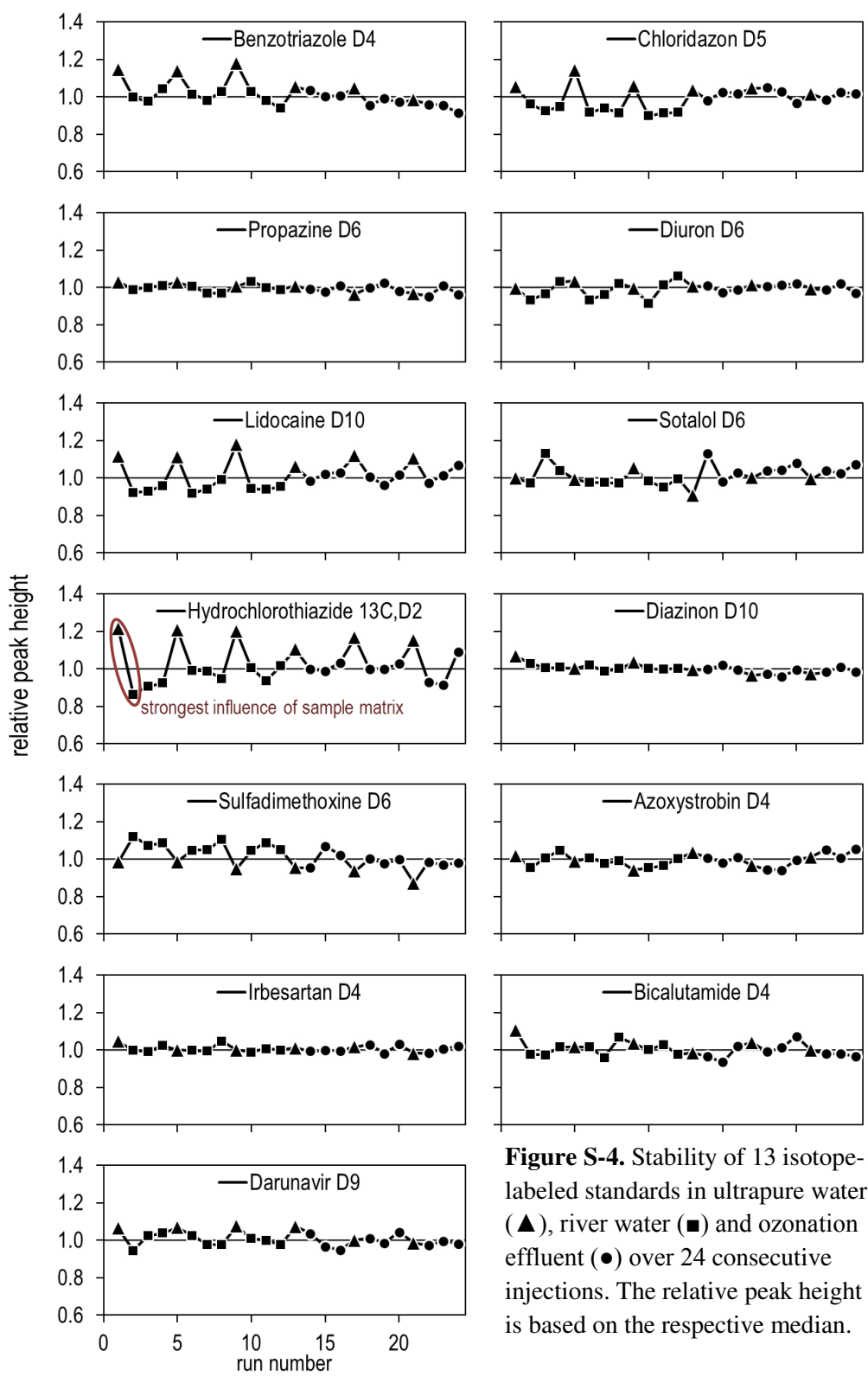
Figure S-4. Stability of 13 isotope-labeled standards in ultrapure water (▲), river water (■) and ozonation effluent (●) over 24 consecutive injections. The relative peak height is based on the respective median.
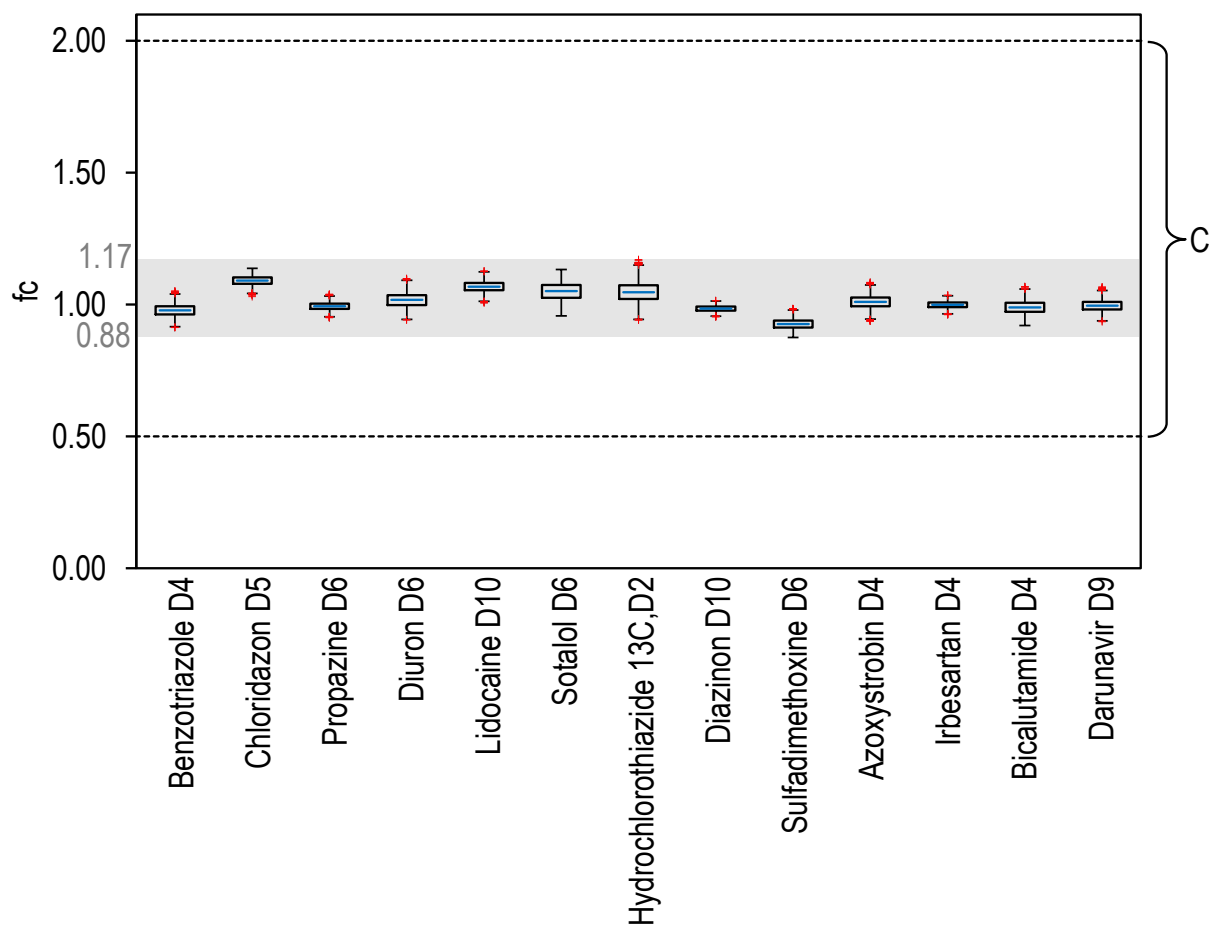
Figure S-5. Boxplots of the fold changes (fc) for the spiked isotope-labeled standards (IS) determined for all pairwise comparisons of triplicates (here: 7056 comparisons for each standard) between the influent (river water) and effluent sample (ozonation effluent) of the process. The intervals for the group *consistency* (C) are indicated by the dashed lines, the minimum ($fc_{min} = 0.88$) and maximum ($fc_{max} = 1.17$) value of the fold changes across all IS are emphasized by the grey rectangle. Boxes: 25th and 75th percentiles (q_1 and q_3), max whisker length: 1.5 (q_3-q_1), blue line: median value, red plus: outlier.

Note that the blank correction was omitted (IS also spiked to blanks) in order to illustrate the system stability and the low impact of matrix effects on the classification.
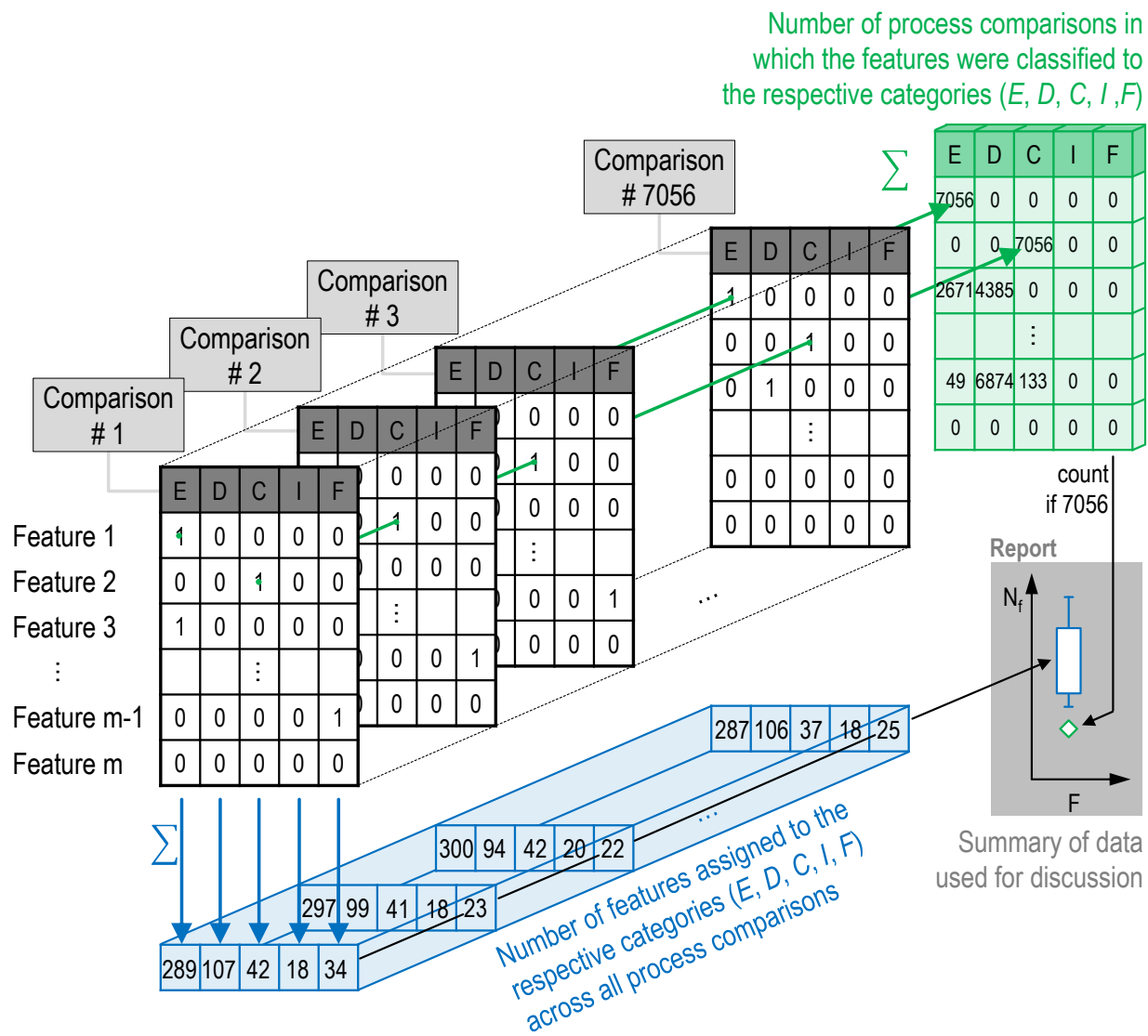
Figure S-6. Data processing strategy for assessing the repeatability of the process comparisons. For each process comparison, each feature was assigned to one respective class (*elimination E*, *decrease D*, *consistency C*, *increase I*, *formation F*) which is indicated by the binary numbers. Blue: sum of all features assigned to the respective category across all comparisons. The distributions of the feature numbers assigned to one certain class (here shown for *F*) were depicted by the boxplots (**Figure 5** in main article). Green: For each single feature, the number of process comparisons (here 7056) were tracked during data processing. Features which were always (across all possible comparison) assigned to one category were counted and additionally considered for the evaluation (green diamond).
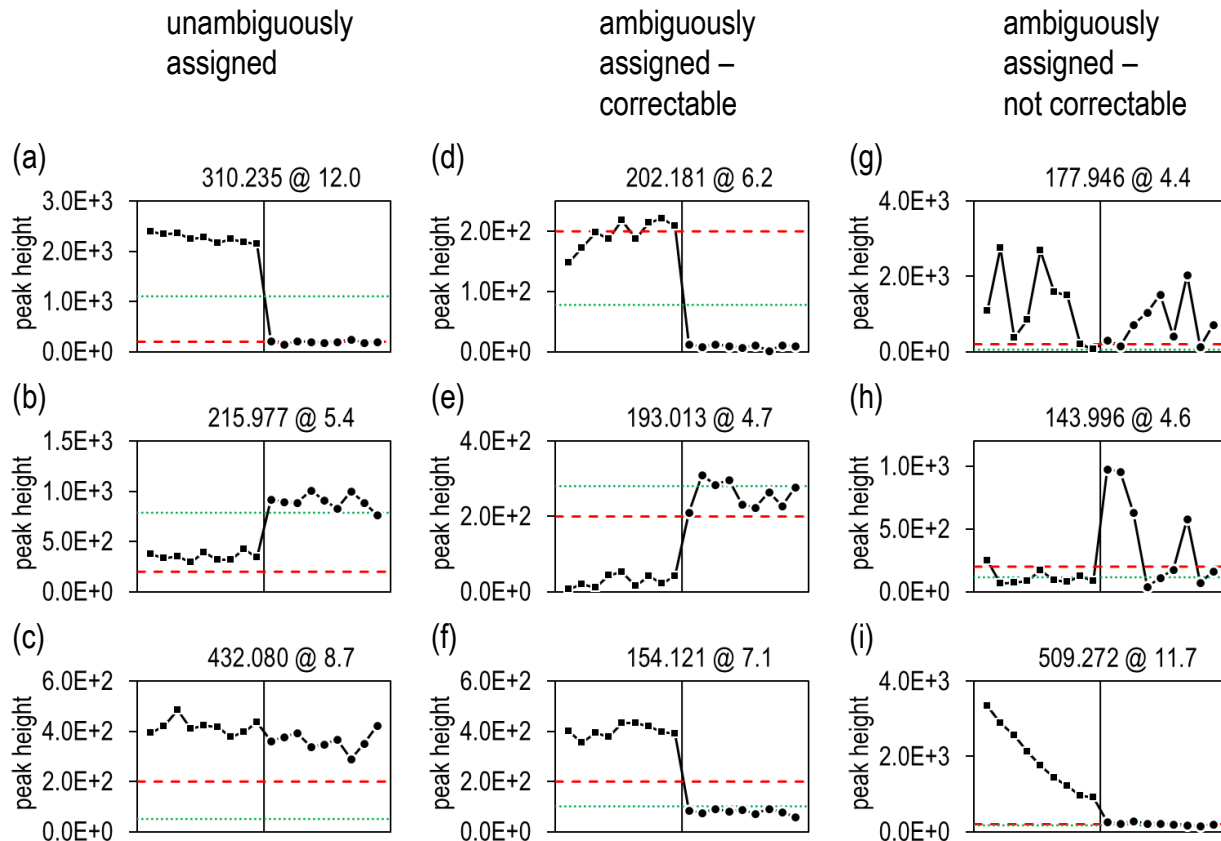
Figure S-7. Selected peak profiles from the ninefold injections of river water (■) and ozonation effluent (●) to illustrate the classification problems using the combinatorial approach. Red dashed line: intensity threshold (200 cps), green dotted line: threshold for blank correction (5 times the blank height).

- Signals unambiguously assigned (in all 7056 comparisons) to:

(a) *elimination*

(b) *increase*

(c) *consistency*

- Signals ambiguously (but correctable) assigned to multiple categories due to:

(d) some influent replicates below / above intensity threshold → feature assigned to *elimination* / *rejection*

(e) some effluent replicates below / above the blank threshold → feature assigned to *formation* / *rejection*

(f) fold change (fc) between sample triplicates of influent and effluent slightly below / above the 0.2 fc limit
    → feature assigned to *elimination* / *decrease*

- Signals ambiguously (and not correctable) assigned to multiple categories due to:

(g, h) implausible intensity profiles of real peaks; some triplicates fulfill all filter criteria and are thus not
        always rejected → feature assigned to multiple categories

(i) trend of peak height during analysis, some triplicates (e.g. Influent #1 #2 #3) fulfill the filtering criteria
    whereas others (e.g. #1 #2 #9) do not → feature assigned to multiple categories.
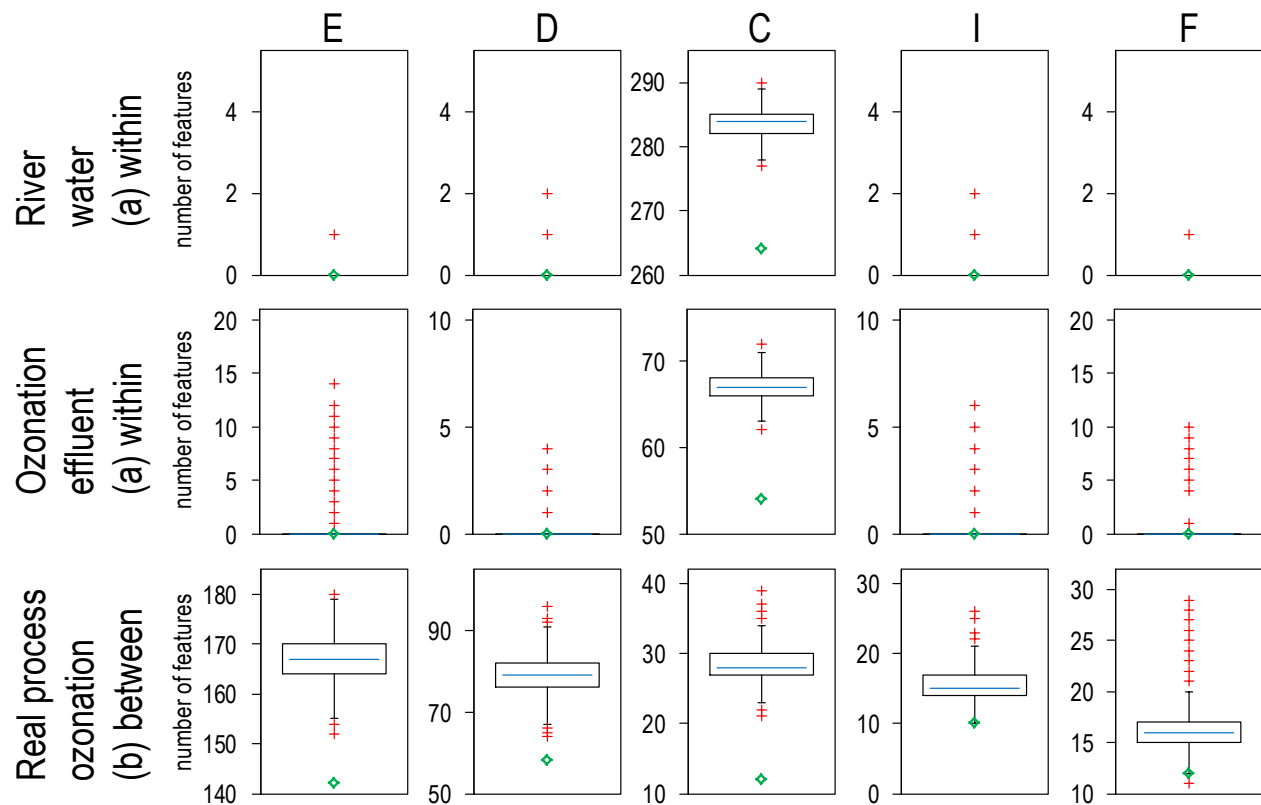
Figure S-8. Boxplots of feature numbers assigned to: *E elimination*, *D decrease*, *C consistency*, *I: increase* or *F formation*. The first and second line shows the within sample comparisons (840 comparisons). The third line shows the between sample comparisons of the real process of ozonation (7056 comparisons). Boxes: 25th and 75th percentiles (q_1 and q_3), max whisker length: 1.5 (q_3-q_1), blue line: median value, red plus: outlier, green diamond: features unambiguously assigned to this group.
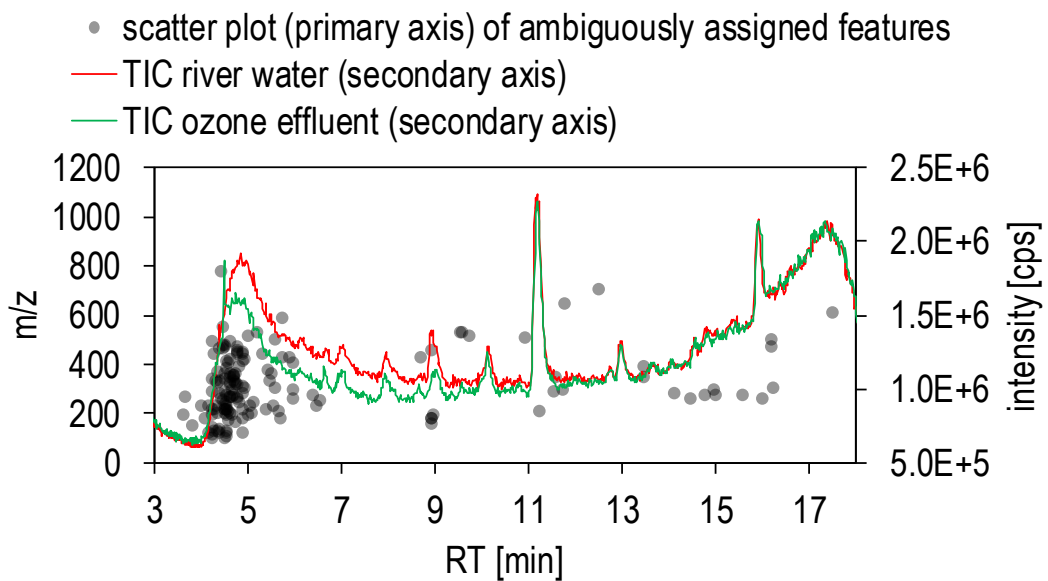
Figure S-9. m/z-RT-scatterplot (primary axis) of all features which are ambiguously assigned to multiple categories and superimposed line plots (secondary axis) of the total ion chromatograms (TIC) of river water and effluent of ozonation.

CHAPTER 1

Application of Non-Target Analysis with LC-HRMS for the Monitoring of Raw and Potable Water: Strategy and Results

Tobias Bader, Wolfgang Schulz, Thomas Lucke, Wolfram Seitz, Rudi Winzenbacher

(**2016**)

Application of Non-Target Analysis with LC-HRMS for the Monitoring of Raw and Potable Water: Strategy and Results

Tobias Bader,[1,2] Wolfgang Schulz,[*,1] Thomas Lucke,[1] Wolfram Seitz,[1] and Rudi Winzenbacher[1]

**[1]Zweckverband Landeswasserversorgung, Laboratory for Operation Control and Research, Am Spitzigen Berg 1, 89129 Langenau, Germany
[2]Sustainable Chemistry and Material Resources, Institute of Sustainable and Environmental Chemistry, Leuphana University of Lüneburg, Scharnhorststraße 1/C13, 21335 Lüneburg, Germany
[*]E-mail: Schulz.w@lw-online.de.**

This contribution focuses on the application of non-target screening by liquid chromatography-high resolution mass spectrometry (LC-HRMS) as a tool in routine water analysis. From the perspective of water suppliers, comprehensive monitoring strategies are required to ensure good water quality. We will illustrate the strengths of the non-target approaches based on three chosen case studies. In principal, the non-target analysis enables monitoring approaches which also cover unknown or unexpected contaminants. The archive function of HRMS data comprises various benefits such as retrospective screening and also allows determining concentrations in a semi-quantitative way. Temporal prioritization in combination with multivariate statistics emerged as helpful strategy for the detection of new contaminants. This is exemplified for the identification of a spill event in river water. The spatial sampling information, on the other hand, allows the localization of possible sources of contamination. Prioritization based on the analytical request reduces thousands of signals to few interesting candidates which increases the success rate during the identification. By such measures, it was possible to prove that a groundwater contamination is caused by an industrial

waste water treatment plant. For the evaluation of processes for water treatment, identification and quantification of individual contaminants is not always of major interest. However, sophisticated strategies are needed to compare the purification efficiency of new technologies to already established ones. Here we evaluated the effectiveness of the fourth treatment step based on activated carbon filtration in waste water treatment plants.

Introduction

The monitoring of drinking water to ensure consistently good quality requires analytical techniques which allow collecting all parameters of interest. The recent developments in high resolution mass spectrometry (HRMS) have initiated new possibilities for the analysis of trace substances without prior knowledge (*1*). This allows detecting contaminants which were not expected or even known before. Comprising high selectivity and sensitivity, the LC-HRMS opens up a new dimension for the monitoring of complex water samples (*2–5*).

However, the identification of unknown contaminants is very time-consuming and, in most cases, requires further analytical techniques (e.g. nuclear magnetic resonance) in order to fully elucidate the chemical structure (*6*, *7*). For several thousands of chromatographic peaks within the data sets, such approaches are practically impossible and thus strategies for prioritization are necessary (*8–10*). If identification is not of major interest, the non-target analysis is still a very valuable tool as it allows comparing different samples (*11*, *12*). Especially for the evaluation of technical processes (i.e. ozonation during water treatment), the non-target analysis offers the possibility of assessing the effectiveness without knowing the identity of the substances which are eliminated or newly formed during a particular process.

In this chapter, we will introduce different approaches for prioritization. Specifically, we addressed three case studies:

- Spill detection in river water (temporal prioritization)
- Localization of sources of contaminations (spatial prioritization)
- Assessing the effectiveness of technical processes (process-related prioritization)

Materials and Methods

Sample Preparation

The sample preparation was kept at a minimum to avoid discriminating against certain substances or substance classes. Samples containing suspended matter were centrifuged for 5 min at 5000 rpm prior to analysis. More details about the sample preparation are listed individually for the different case studies (section *background and objective*).

LC-MS Sample Acquisition

The samples were analyzed with a Shimadzu (Kyoto, Japan) LC20 series HPLC system coupled to the Sciex (Framingham, USA) QTOF/MS TripleTOF™ 5600 equipped with a DuoSpray™ Ion Source. In two separate runs using the same chromatography, positive and negative electrospray ionization was used for the generation of ions. After a survey scan (m/z 100 - 1200), a maximum number of 12 data dependent acquisition (DDA) experiments was performed within one single cycle. The acquisition of the MS² spectra within a mass range of m/z 30 - 1200 was accomplished with a collision energy ramped from 25 to 55 eV.

An Agilent (Santa Clara, USA) Zorbax Eclipse Plus C18 (2.1 x 150 mm, 3.5 μm) column was connected to a Phenomenex (Aschaffenburg, Germany) precolumn Security Guard AQ C18 (4 x 2 mm). During the runs, both columns were maintained at a temperature of 40 °C. The flow rate was kept at 300 μL min⁻¹ with a water (A) and acetonitrile (B) gradient containing 0.1% (v/v) formic acid, respectively. A direct injection of 100 μL of each sample was accomplished without any preconcentration (such as solid phase extraction). A blank value was generated by a zero-injection (i.e. injection of 0 μL) in order to cover contaminations caused by the system itself. Detected contaminations were excluded from DDA throughout the complete chromatographic run.

Software Tools

Peak finding and alignment were performed with MarkerView™ (1.2.1). The quantitation package MultiQuant™ (3.0.2) was used for subsequent peak integration. Both software tools are commercially available from Sciex. Componentization was accomplished by a self-written script in Matlab (R2015a, MathWorks, USA), where further filtering steps were included.

Data Processing

The complete workflow including the main steps of the data processing is depicted in Figure 1. The sampling was designed based on the objective of the respective study. All samples were analyzed using the same LC-HRMS method. In the first steps, a peak finding algorithm extracts all peaks exceeding a given noise threshold (device and sample specific) from the raw data. Such peaks which are characterized by their mass-to-charge ratio, retention time and intensity are referred to as *features*.

The results are set on so called peak lists. During the peak alignment, the same features from different samples are merged - based on mass and retention time tolerances (here: 10 ppm and 0.15 min). The peaks table is the result of the merged peak lists and represents a matrix where the detected features (m/z, RT) represent the variables (lines) while the objects (columns) are the different samples. The matrix entry is the response of a certain feature in the respective sample.
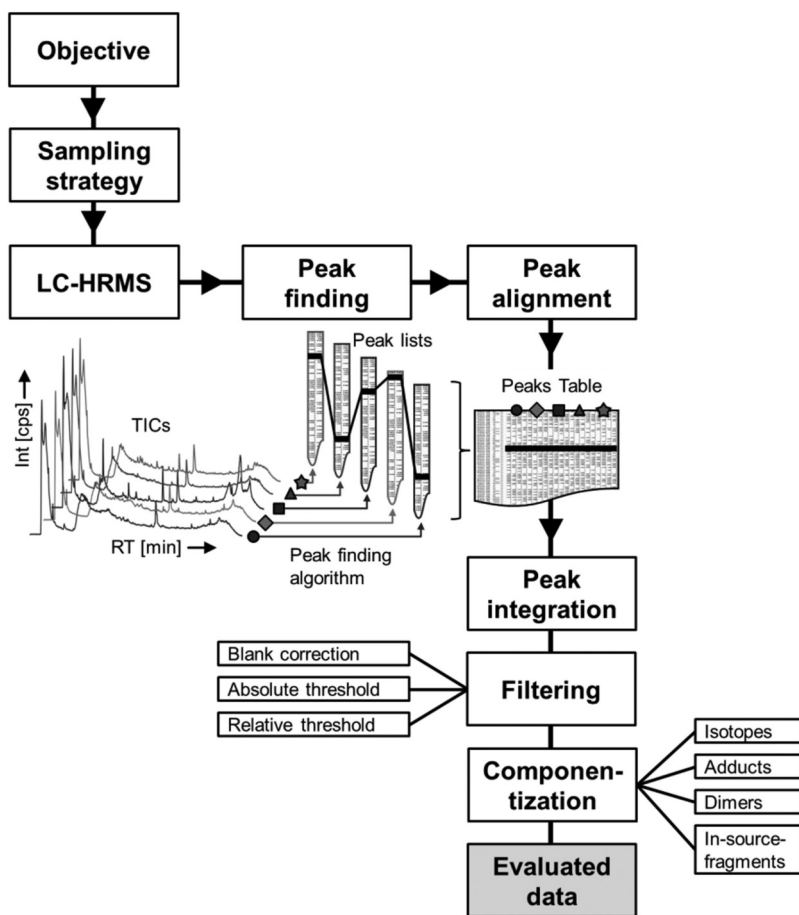
Figure 1. Workflow of the data preprocessing and filtering of features for a specific objective.

To reduce the problematics of false positive and false negative findings, a subsequent peak integration step (recursive) was used to preserve more peak characteristics (area, height, width and retention time) which can easily be used for further filtering steps. These allow reducing false positives (i.e. the peak finding algorithm annotated a background / matrix signal as a peak) while the recursive integration of sample replicates allows for correcting partially false negative findings (i.e. the peak finding algorithm missed a real peak of interest in some of the replicates) (*13*). The first filtering step is the blank correction where all features whose peak area and peak height does not significantly differ from the blank's ones are eliminated. The exact parameters which are used for

the filtering steps can be taken from Table 1. During the absolute thresholding further criteria must be satisfied. Each feature has to exceed a certain peak area and peak height. Based on the used chromatography, retention times within 2 - 19 min were extracted (log P ≈ -1 to 5). Furthermore, a certain range for the peak width at full width at half maximum (FWHM) was defined. All filter criteria were optimized using more than 200 different reference standards at different concentration levels (data not shown).

Each sample was analyzed in duplicate or in triplicate, which allows assessing the fluctuation of a certain peak characteristic. The LC-HRMS system has proven to be very stable, which is why the fluctuation of the peak characteristics can be applied as a further filter criteria (*relative threshold*). This was accomplished by forming the quotients for characteristic (peak area, peak height, peak width and the retention time) between the technical replicates. If the calculated quotient is out of range, meaning that the fluctuation cannot be explained by measurement uncertainty, the corresponding features are rejected.

Table 1. Parameter for Filtering of the Features

	Minimum value	*Maximum value*
Blank correction		
Fold change (f)	5	∞
Absolute threshold		
Peak area (A)	350 cts	∞ cts
Peak height (H)	100 cps	∞ cps
FWHM[*] (W)	3.5 sec	12 sec
Retention time (RT)	2 min	19 min
Relative threshold		
Quotient $\frac{x_1}{x_2}$ with x = A, H, W	$\frac{2}{3}$	$\frac{3}{2}$
Quotient $\frac{RT_1}{RT_2}$	$\frac{99}{100}$	$\frac{100}{99}$

[*]: Full width at half maximum

The final matrix contains all features and samples but only intensity values different from zero for those features which have met all filtering criteria (blank correction, absolute and relative thresholding).

Before the data set is used for further analysis, the so called componentization was performed. During this step, isotopes, adducts, dimers and common in-source-fragments (such as loss of H_2O or CO_2) are assigned to a potential [M+H]$^+$ or [M-H]$^-$ ion. However, the different ion species are not removed from the existing matrix but rather annotated and also used for further steps. All dataset which are described in this chapter have been processed with the workflow as mentioned above.

Results and Discussion

The application possibilities of non-target screening in "real-world-examples" are illustrated using three different case studies. These examples where chosen to cover a wide range of prioritization strategies. A summary of the case studies is illustrated in Figure 2 and described in more detail below.
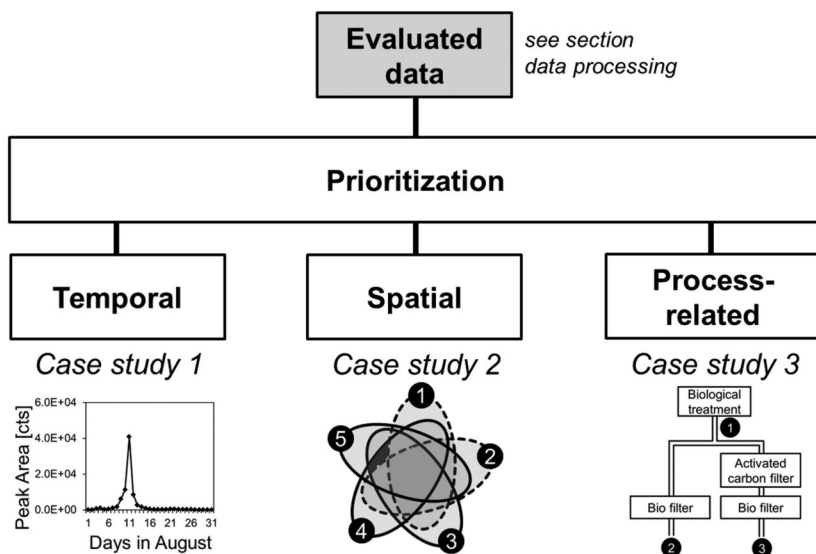

Figure 2. Overview of case studies described in this chapter.

Case Study 1 - Non-Target Analysis for the Monitoring of Water Resources

In this specific case study, river water was directly used (without bank filtration) for the production of potable water. Many contamination sources (e.g. agriculture, household, industry) may influence the river water and may therefore affect the quality of the drinking water. In this example, we illustrate a spill detection in the river water which was discovered by means of non-target analysis.

Case Study 2 - Non-Target Analysis as a Forensic Tool in Water Analysis

A groundwater contamination was already observed after target analysis. However, different sources of contamination were theoretically possible. Two municipal and one industrial waste water treatment plant (WWTP) were suspected to be responsible for the contamination of the groundwater site. The aim of this approach was to show that the industrial WWTP has a major impact on the groundwater sample.

Case Study 3 - Non-Target Analysis To Assess Technical Processes

Continuous modernization of water treatment technologies requires analytical tools in order to assess their performance and to compare them with already established technologies. Most conventional trace analytical techniques do not allow detecting unknown compounds such as transformation products (TPs). For the evaluation of technical processes, however, TPs are of major interest. In this example we illustrate the use of non-target analysis to assess activated carbon filtration as a fourth treatment step, i.e. an additional step after biological treatment, and compare it to the conventional treatment of waste water.

Case Study 1 - Non-Target Analysis for the Monitoring of Water Resources

Background and Objective

Routine monitoring of raw and potable water is essential to verify the quality and to control the water treatment processes in the waterworks. In addition to targeted programs, non-target analysis is a valuable tool to also cover unknown or unexpected compounds. This case study focuses on a waterworks where river water (#1) is directly used (without bank filtration) for the production of potable water (#2) as is exemplified in Figure 3. Contaminations in the river water might affect the quality of the drinking water. This circumstance emphasizes the importance of monitoring programs.

On each day of the year 2014, a 24-h composite sample was taken from the river water as well as from the potable water. In a first step, only one random sample per week was analyzed, while the other samples were stored as retention samples and only analyzed if the first data evaluation revealed any abnormalities.

Especially unforeseeable spill events, i.e. substances occurring in high concentrations within a short period of time, may possibly represent challenges for the water purification process. The aim of this study was to detect such spill events and proof whether the quality of the potable water has been affected.

Filtering Processed Data for Spill Detection

After processing, the resulting data set was screened for spill events. To accomplish this, features whose maximum intensities were much higher than their 80% quantile were extracted from the existing data set.

A prior period of 25 calendar weeks was defined to get more robust statistics. The extracted features were afterwards ranked in a way that highest intensities and lowest frequencies will result in the highest scoring. One of the best scores was obtained for the feature with m/z 337.213 at 5.4 min. After recognizing a possible spill in calendar week 33 (Figure 4a), all daily taken retention samples were analyzed at each day of August 2014 (Figure 4b).
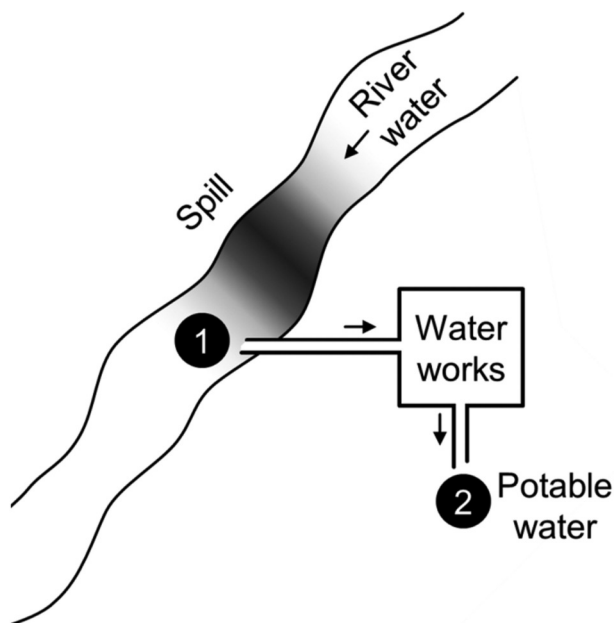
Figure 3. Schematic illustration of the sampling sites, daily retention samples from the river water sample (#1) and the potable water sample (#2) were taken and one sample randomly analyzed per week.
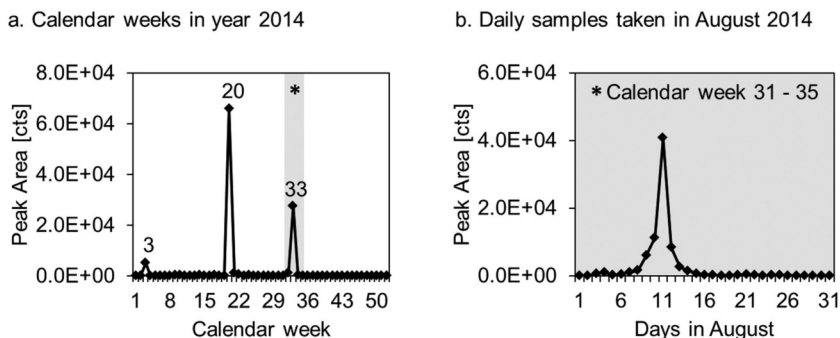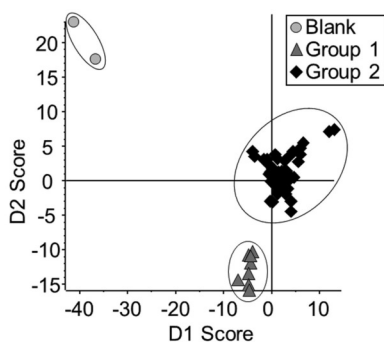

Figure 4. Time series of an unknown compound (m/z 337.213 at 5.4 min) in the river water; time profile for a. the year 2014 (one sample at each calendar week) and b. August 2014 (one sample per day).

The profile of the daily samples shows that the spill event covers a period of about four days (Aug 9th to Aug 12th). The maximum intensity was reached on August 11th. However, the potable water did not show any signal for this compound. Even in calendar week 20, where higher signal intensities where observed, the signal was absent after the treatment steps. This leads to the conclusion that the substance has been removed / transformed (below the limit of detection) during the water treatment in the waterworks.

Identification

Using the accurate mass, the isotope pattern and the product ion spectrum, the vendor software was used to calculate the elemental composition. The best score in the MS as well as in the MS² was obtained for $C_{18}H_{28}N_2O_4$. For the identification, different databases were searched against the generated formula. The Drugbank (*14*) database suggested the beta blocker acebutolol as a possible candidate. The acquired MS² spectrum showed a perfect match with the literature spectrum documented in MassBank (*15*) which has prompted to purchase a reference standard. Finally, the proposed candidate was identified and also quantified in the river water samples. In August 2014, a maximum concentration of about 60 ng L^{-1} was reached. Regarding the complete spill event in August, the cumulative load of acebutolol was about 1 kg. The highest concentration in the year 2014 was about 100 ng L^{-1} and could be detected in calendar week 20 (see Figure 4a).
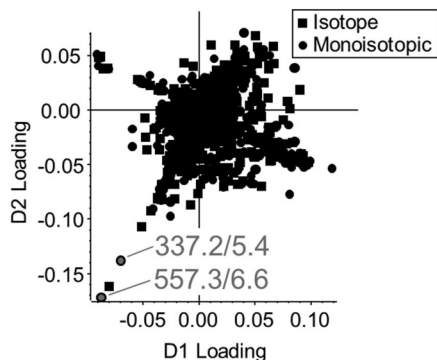
Figure 5. Similarity search with principal component analysis and discriminant analysis (PCA-DA); a. Scores plot after grouping blanks (circles), river samples from August 9th to August 12th (Group 1, triangles) and the remaining river samples from August (Group 2, diamonds); b. Loadings plot with discriminating features highlighted in the lower left corner.

Searching for other features that follow a similar profile was the next logical step. It is likely that other compounds may show a similar time profile and hence may be in conjunction with the detected spill event. For similarity-search within huge data sets, multivariate statistics such as cluster analysis or principal component analysis represent helpful tools as it is possible to rapidly structure big matrices and retrieve the information of interest. In this particular study, we have implemented a combination of principal component analysis and discriminant analysis (PCA-DA) to find features with similar time profiles to acebutolol. Knowing that spill event occurred between August 9th and Aug 12th, the groups for the DA were defined. One group comprises the dates of the spill detection, the other group all remaining samples taken in August (except Aug 9th - 12th). A third group covering two zero-injections was defined. The scores and loadings plot of the PCA-DA approach are illustrated in Figure 5.

As is evident, the two groups were nicely separated by the PCA-DA approach. The variables, hence features, which are responsible for the separation are scattered in the loadings plots (Figure 5b). Features which are highly responsible for the separation of Group 1 and Group 2 are located in the lower left corner (quadrant III). As expected, acebutolol (highlighted 337.2/5.4) has a high impact on the separation of group 1 and 2. The monoisotopic mass as well as the isotope of a second feature also possess a high load (highlighted 557.3/6.6). The accurate mass of m/z 557.255 was detected at a retention time of 6.6 min. The profile of acebutolol and the unknown feature are plotted in Figure 6.
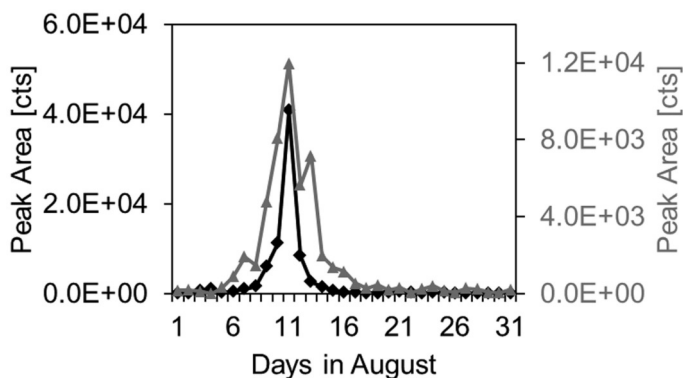

Figure 6. Superimposed time series profiles of acebutolol (black diamonds, primary axis) and the unknown feature (m/z 557.255 at 6.6, grey triangles, secondary axis) in August 2014.

The time profiles show a good fit which underlines the success of the similarity search by PCA-DA. Interestingly the unknown feature did not occur in calendar week 20, where high loads for acebutolol could be observed (see Figure 4a). The unknown compound was searched against most common spectral libraries. In-silico fragmentation using Metfrag (*16*) could only explain one fragment ion and did not provide any further information. The identification remained at level 5 (*17*) as it was not even possible to assign a unique elemental

composition. This example shows how difficult any further steps are, if no MS² information is available in public databases. It should be noted that the accurate fragment masses and expertise knowledge on fragmentation rules might advance on the tentative identification.

Case Study Conclusion

The spill detection in the river water emphasizes the need for broader screening techniques. Non-target screening for the monitoring of raw and processed water represents a very powerful monitoring tool which also allows for the detection of unknown contaminations. Besides acebutolol, another spill caused by the vasoactive drug buflomedil was detected in the past. These spill events might indicate an impact of the pharmaceutical industry on the river water. In future studies, the exact locations of the discharge should be detected.

These examples, however, only show the results for contaminants that could be fully identified. The identification of the remaining list of "interesting" courses represents a main challenge in the future and will not always lead to a successful identification. From the water supplier's perspective, features passing through the process of water treatment are of major interest. This prioritization significantly reduces the list of candidates. However, special focus should also be given to transformation products which were formed during the treatment steps (e.g. ozonation). Again, the time profiles and the application of multivariate statistics offer a valuable tool to correlate the formation of transformation products with occurrence of certain contaminants in the river water.

Case Study 2 - Non-Target Analysis as a Forensic Tool in Water Analysis

Background and Objective

During routine target monitoring programs, findings of elevated levels for various chemicals clearly indicated the anthropogenic contamination of a groundwater sample. Besides municipal WWTPs, also the chemical industry is located near the groundwater monitoring well. It has been suggested, that the industrial WWTP might (also) be responsible for the detected contamination. A schematic map of the sampling sites is given in Figure 7. Two municipal (#1 - 2) and an industrial (#3) WWTP discharge into the nearby river (#4). The contamination in the groundwater (#5) might be explained by infiltration of the river water.

The aim of this particular study was to prove that the industrial WWTP is (also) responsible for the contamination of the groundwater. In other words, is it possible to find contaminants which can exclusively be attributed to the industrial WWTP?
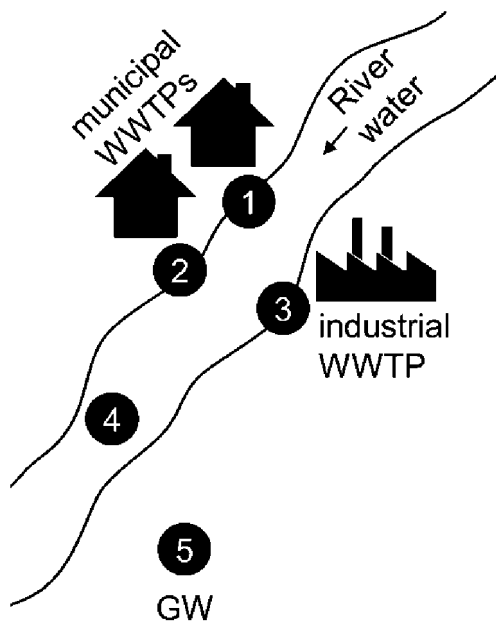
Figure 7. Schematic illustration of the sampling sites. Two municipal WWTPs (#1 and #2) as well as an industrial WWTP (#3) discharge into the river (#4). Via bank filtration, the river water infiltrates into the groundwater (#5).

Prioritization of Processed Data

Grab samples were taken from all locations (#1 - #5) and analyzed by LC-HRMS. As a first step, a suspect-target list (covering 3200 substances) was queried with the objective to find such "source-specific" contaminants which would allow to show the industrial impact on the groundwater well. However, none of the detected substances was specific for the industrial WWTP (i.e. did not occur in the two municipal WWTPs) while also been detected in the groundwater monitoring well (#5). As a consequence, it was not possible to answer the analytical request based on a suspect-screening approach. The same data set has then been used for processing based on a non-target approach. After peak finding and alignment, between 4300 and 6700 peaks were detected in each single sample. By additional filter criteria (as has been described), the numbers were reduced to about 50% and after componentization between 1400 and 2400 peaks remained.

For further prioritization of candidates of interest, filter criteria based on the analytical request were designed. This has been accomplished by simply applying logical connections between the different samples (*12*). Only features which were present in the industrial WWTP (#3), the river water (#4), the contaminated groundwater (#5) but not in the two municipal WWTPs (#1 - 2) were extracted from the data set. By this measure, the thousands of peaks could be reduced to 54 candidates that might help answering the initial question. In the last step,

the identification of some promising candidates was the objective as robust and reliable results were needed for further actions.

Identification of Contaminants

All samples were acquired using data dependent acquisition. For 33 (out of 54) precursor ions, the DDA approach has automatically generated useable MS² spectra. The accurate masses of the 33 features were uploaded to the platform FOR-IDENT (*18*) which represents a compilation of water relevant chemicals. In the next step, all possible candidates which matched the mass of a feature (+/- 10 ppm) were linked to the corresponding acquired MS² spectrum. All MS² spectra were automatically exported from the vendor software and via FOR-IDENT automatically sent to the in-silico-fragmentation tool Metfrag (*16*). Seven out of 33 candidates were ranked with high Metfrag scores.
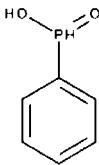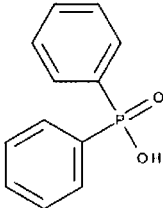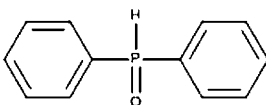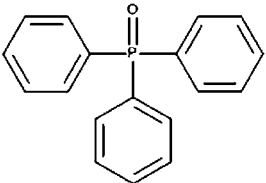
One compound was finally identified and quantified using a reference standard (level 1) (*17*). Three out of seven compounds were tentatively identified (level 2a) by comparing the MS² spectra with literature spectra found in MassBank (*15*) and/or mzCloud (*19*). The three remaining candidates represent level 2b identifications, i.e. structural evidence based on MS² information, but no literature information available/found. Six out of seven hits representing industrial chemicals as well as transformation products. Four of these contaminants are listed in Table 2. The two remaining candidates are most likely hydroxylated forms of triphenylphosphine oxid (TPPO) but could not be confirmed by literature spectra (level 2b). For the final proof of the proposed substances, the reference standards have to be purchased. However, the occurrence of several substances from similar substance classes increases the confidence of the results. In addition to those substances, the pharmaceutical Praziquantel (the seventh hit) was also identified at level 2a.

The extracted ion chromatograms of the four phosphorous contaminants are illustrated in Figure 8. The signal intensity in the industrial WWTP was set to 100% the one in the groundwater is relatively plotted based on the WWTP. The signal in the groundwater sample lies within 2% for diphenylphosphine oxide (DPPO) and 380% for diphenylphosphinic acid (DPPA). Please note that these grab samples only represent a snapshot of the entire picture, meaning that sampling on another day may possible lead to other concentrations especially in the industrial WWTP and therefore changes the relative differences. Nevertheless, these finding clearly indicate the input of the pollutants into the groundwater.

Except phenylphosphinic acid (PPA), all pollutants were already described in the literature (*20*). The contamination with TPPO is much likely related to the Wittig reaction (*21*) where it is formed as a byproduct (*22*). It is notable that the concentration of TPPO in the contaminated groundwater sample was higher than 5 µg L^{-1}. The other phosphorous compounds listed represent very likely degradation products as has already been described for DPPO and DPPA (*20*).

As might be expected, PPA and DPPA could also be detected in the negative ionization mode. None of these contaminants were detected in the municipal WWTPs.

Table 2. Substances (Tentatively) Identified in the Industrial WWTP and in the Groundwater

Name	Formula	Structure	Level of identification
Phenyl-phosphinic acid (PPA)	C6H7O2P		2b
Diphenyl-phosphinic acid (DPPA)	C12H11O2P		2a
Diphenyl-phosphine oxide (DPPO)	C12H11OP		2a
Triphenyl-phosphine oxid (TPPO)	C18H15OP		1

Case Study Conclusion

Based on the analytical request, thousands of peaks could be reduced to less than 60 specific peaks, i.e. these signals were not recognized in the municipal WWTPs but occurred in the industrial one, the river water as well as in the groundwater. The in-silico fragmentation was very helpful to rank the candidate structures and fastened the whole procedure as only the high scoring structures were compared with spectral libraries. It was possible to tentatively identify six out of 54 while TPPO could be fully identified and quantified.

By using a non-target approach, the initial question could successfully be answered: The industrial WWTP has an impact on the regarded groundwater site which could be shown by source-specific contaminants such as PPA, DPPA, DPPO and TPPO. The contamination caused by TPPO could be fully identified and also quantified by LC-HRMS.
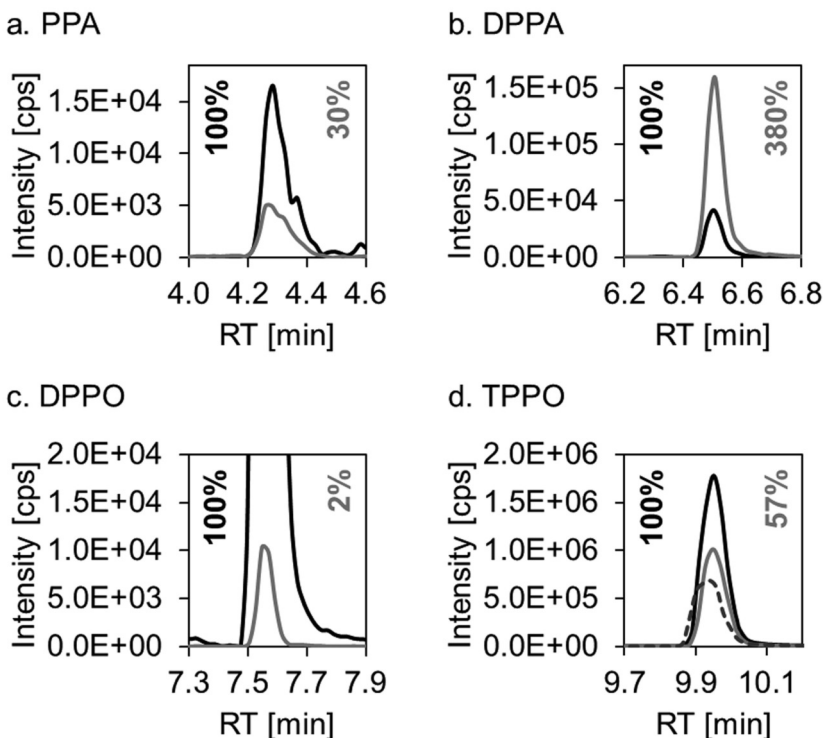

Figure 8. Extracted ion chromatograms for PPA (a.), DPPA (b.), DPPO (c.) and TPPO (d.) in the industrial WWTP (black trace) and in the contaminated groundwater sample (grey trace). In case of TPPO, a 5 µg L⁻¹ standard is shown by a dashed line.

Whether or not the suspect-screening approach will be successful depends on the list which is used for screening. Our list comprising more than 3200 components did not reveal source-specific contaminants. The non-target approach, on the other hand, does not require any prior information and was therefore expedient (though more complex) for this particular request.

Background and Objective

   This study was aimed at assessing the purification efficiency of the fourth treatment step of a waste water treatment plant (WWTP) in comparison to the conventional waste water treatment. Therefore, one sample was taken after the secondary clarifier (#1) before the stream was divided into the two branches, meaning that both processes were driven with the same influent (Figure 9). The left branch represents the conventional waste water treatment the right one comprises the activated carbon filter. Samples were taken after the bio filters of the two different processes (#2 and #3). The samples in this study were 24 h-composite samples filtered by a 0.45 µm membrane filter.
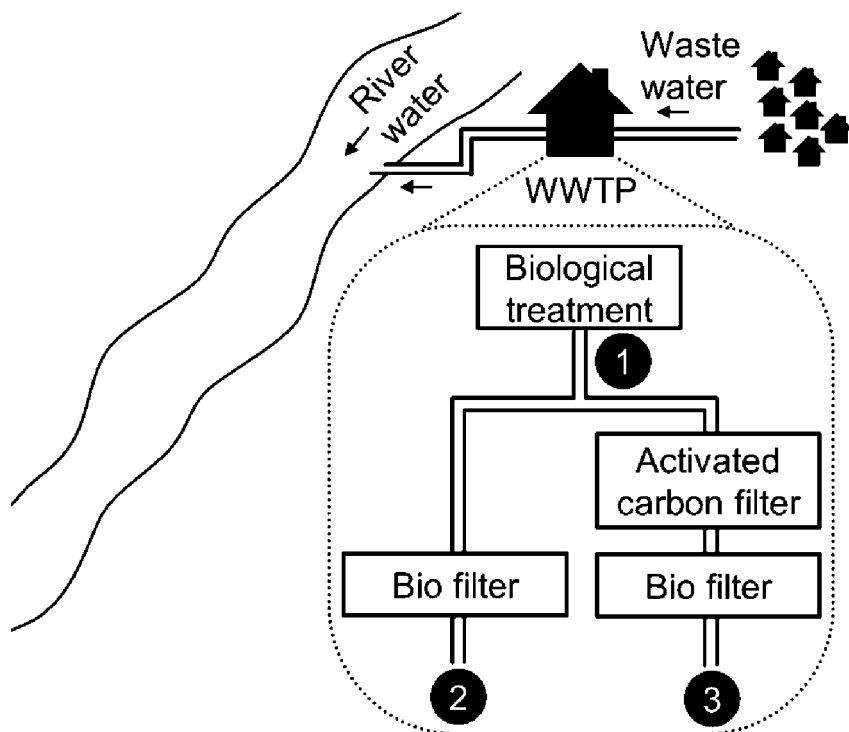

Figure 9. Sampling points for the comparison of conventional waste water treatment (biological treatment, left branch, #1 → #2) and treatment with a fourth purification step using activated carbon filtration (right branch, #1 → #3).

Comparative Evaluation

After data processing (see section *data processing*) the obtained feature lists were compared to investigate the difference between the conventional waste water treatment and the additional filtering step using granulated activation carbon (GAC) filtration. It was expected that the additional filtering step is an added value with regard to the cleaning performance.

By comparing the two process branches, the adsorption efficiency of the GAC filtration should be considered. The comparison of sample #1 and #2 covers the way of the conventional treatment. For the treatment including the adsorption step, sample #1 and #3 were compared. For assessment of the efficiency, the influent and effluent samples of the respective branch were used for classification of the detected features. The groups were defined as follows:

- Elimination: only detected in influent
- Formation: only detected in effluent
- Common: detected in influent and effluent

At this point, only a summary inspection of all features was conducted. The identification or even quantification of individual candidates was not within the scope of the case study.

This categorization was conducted for both individual process branches. The results are summarized in Figure 10. Comparing the 6602 features in the influent sample (#1) with the two different effluents, the proportion of common features, i.e. features detected in both the influent and effluent samples can be derived. For the conventional treatment, 4952 features were common, the fourth treatment step only revealed 4264 common features. The higher purification performance using GAC filtration can already be assumed by comparing the number of features in the two effluent samples (#2 and #3). For the conventional treatment 6183 features were detected, while the detected 5308 features using the advance purification represents a reduction of more than 14%. Furthermore, the number of eliminated features is about 42% higher (1650 vs. 2338) if using GAC filtration. On the other hand, the conventional treatment reveals about 15% more newly formed features (1044 vs. 1231). To sum up, the additional treatment step including GAC filtration increases the percentage of elimination and decreases the amount of newly formed compounds.

However, this simplified approach does not consider changes in the relative intensities of the common features, regardless of how high they might be. To also take this aspect into account, a closer look was taken at the features detected in both influent and effluent samples (Figure 10, detailed view). By comparing the relative changes of the signal intensities, the common features were divided into different groups. Features subjected to elimination (i.e. features with decreasing intensities throughout the process) were divided into three sub classes:

- Elimination of less than 20%
- Elimination between 20% and less than 60%
- Elimination of at least 60%

Features with higher signal intensities in the effluent sample were categorized in two different groups:

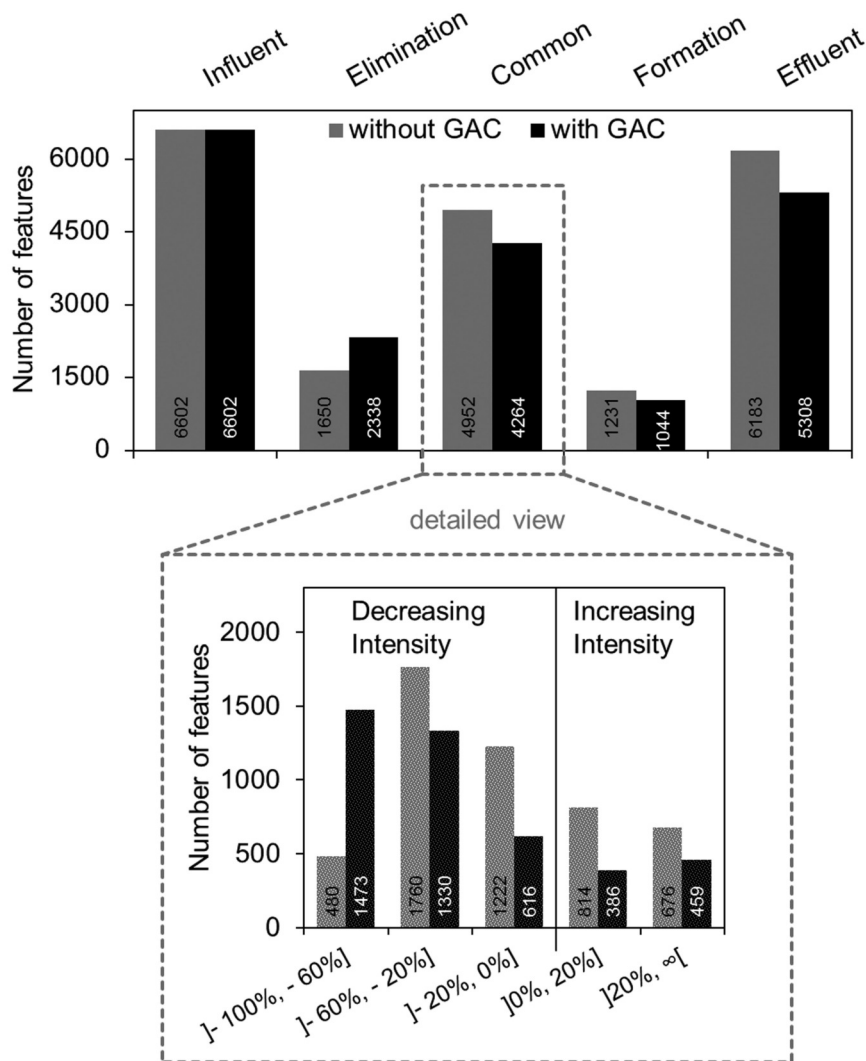- Increase of up to 20%
- Increase of more than 20%

Figure 10. Comparison of the waste water treatment with (black bars) and without (grey bars) granulated activated carbon (GAC) filtration.

The distribution of features assigned to the different categories is illustrated in the *detailed view* of Figure 10. As is evident, the proportion of features eliminated by at least 60% are about three times higher if comparing the fourth treatment step with the conventional one (480 vs. 1473). The positive effect of the GAC step seems to be significant. Moreover, the fraction of features with increasing intensities troughout the process is more than 75% higher for the conventional treatment without the GAC filtration (814 + 676 vs. 386 + 459).

Matrix effects were expected to only have minor influence on the results as the complex influent sample was the same for both process branches. The comparative evaluation is less prone to misinterpretations as both processes are subjected to similar matrices.

Please note that especially for such evaluations the reliability of the complete workflow (see section *data processing*) needs to be very high to avoid skewing of results. High percentages of randomly occurring false positive and negative findings highly hamper the assessment and therefore have to be kept at minimum. To get an impression of the significance of the results, single features were selected on a random basis and manually reviewed. This step reveals only few false positive and false negative (i.e. real peak filtered during the processing) findings. Furthermore, the complete assessment was done for three individual sample series each in positive and negative ionization mode (only one series with positive ionization is shown here). The results across the three sampling series were largely consistent suggesting that randomly occurring errors are clearly underrepresented.

Case Study Conclusion

The additional benefit of the fourth treatment step could be shown by comparing samples using a non-target approach. The simple comparison of influent and effluent samples clearly indicated a lower rate of newly formed features and at the same time a higher elimination for the additional filtering step. A closer look at the common features also revealed a better rate of elimination compared to conventional treatment.

The process evaluation without the primary objective of identifying features reduces the efforts. However, the process comparison can also be seen as a prioritization tool, as for example only transformation products (i.e. newly formed compounds) are of interest.

The applied strategies are transferable to other technical processes such as ozonation. Contrary to most common techniques, the non-target analysis uses all detectable substances for the process comparison. This represents a major benefit for the assessment of processes as other approaches might have difficulties with regard to newly formed compounds. Assuming that a large amount of transformation products is not known yet, non-target screening allows a very extensive evaluation of new technologies and treatment steps.

Summary and Conclusion

The three case studies clearly show the strengths of non-target analysis for wide-ranging different applications. Prioritization is a key factor, as - in most cases - few signals already allow to answer the analytical request. Multivariate statistics are helpful to extract these candidates from the wealth of data.

The LC-HRMS opens up a new dimension for the monitoring of water resources. Covering unknown and unexpected compounds, non-target screening provides a more complete picture of the fate of trace organic contaminants in the aquatic environment but also during the different processes of water treatment. Fast processing tools for data handling, just-in-time analysis and well thought prioritization tools will support such complex approaches in the future and might allow an almost simultaneous derivation of measures.

In most instances, the subsequent approach is to localize the source of the contamination which would allow the derivation of countermeasures for the water protection. Especially short arising spill events in rivers covering only a few days represent a main challenge for the identification of possible polluters. A denser sampling network should be established to narrow down the possible territory.

For the comparison and evaluation of the effectiveness of new technologies in water treatment, non-target approaches will be important in the near future. As the final identification of unknown compounds is challenging and requires, in most cases, further analytical techniques, approaches where the identification is not of major interest do only have a very limited number of drawbacks. However, if regarding processes based on the number of features, special attention should be given to the repeatability of the results. Therefore, extensive method development and sophisticated data processing workflows are indispensable to obtain reliable results.

References

1. Krauss, M.; Singer, H.; Hollender, J. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem* **2010**, *397*, 943–951.
2. Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* **2014**, *184*, 25–32.
3. Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* **2015**, *49*, 12333–12341.
4. Schymanski, E. L.; Singer, H. P.; Longrée, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripollés Vidal, C.; Hollender, J. Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2013**, *48*, 1811–1818.
5. Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M.;

Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **2015**, *407*, 6237–6255.

6. Wendel, F. M.; Lütke Eversloh, C.; Machek, E. J.; Duirk, S. E.; Plewa, M. J.; Richardson, S. D.; Ternes, T. A. Transformation of Iopamidol during Chlorination. *Environ. Sci. Technol.* **2014**, *48*, 12689–12697.

7. Brezina, E.; Prasse, C.; Wagner, M.; Ternes, T. A. Why Small Differences Matter: Elucidation of the Mechanisms Underlying the Transformation of 2OH- and 3OH-Carbamazepine in Contact with Sand Filter Material. *Environ. Sci. Technol.* **2015**, *49*, 10449–10456.

8. Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P. 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.* **2015**, *87*, 145–154.

9. Letzel, T.; Bayer, A.; Schulz, W.; Heermann, A.; Lucke, T.; Greco, G.; Grosse, S.; Schüssler, W.; Sengl, M.; Letzel, M. LC–MS screening techniques for wastewater analysis and analytical data handling strategies: Sartans and their transformation products as an example. *Chemosphere* **2015**, *137*, 198–206.

10. Schollée, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* **2015**, *87*, 12121–12129.

11. Nürenberg, G.; Schulz, M.; Kunkel, U.; Ternes, T. A. 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* **2015**, *1426*, 77–90.

12. Müller, A.; Schulz, W.; Ruck, W. K. L.; Weber, W. H. A new approach to data evaluation in the non-target screening of organic trace substances in water analysis. *Chemosphere* **2011**, *85*, 1211–1219.

13. Bader, T.; Schulz, W.; Kümmerer, K.; Winzenbacher, R., General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Anal. Chim. Acta.* http://dx.doi.org/10.1016/j.aca.2016.06.030.

14. Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–72.

15. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: a public repository

for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.

16. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminf.* **2016**, *8*, 1–16.

17. Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098.

18. *FOR-IDENT: Improvement in the identification of organic trace substances: Merging of resources and standardization of suspected- and non-target analysis*; https://water.for-ident.org/login (accessed 01 Dec 2015 (login only)).

19. *HighChem mzCloud: Advanced Mass Spectral Database*; www.mzcloud.org (accessed 15 Apr 2016).

20. Knepper, T. P.; Karrenbrock, F. The Rhine; Knepper, T. P., Ed.; In *The Handbook of Environmental Chemistry 5*; Springer: Berlin, 2006; Vol. Part L, pp 235−254.

21. Maryanoff, B. E.; Reitz, A. B. The Wittig olefination reaction and modifications involving phosphoryl-stabilized carbanions. Stereochemistry, mechanism, and selected synthetic aspects. *Chem. Rev.* **1989**, *89*, 863–927.

22. Schlüsener, M. P.; Kunkel, U.; Ternes, T. A. Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants. *Environ. Sci. Technol.* **2015**, *49*, 14282–14291.