NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES

DEPARTMENT OF CHEMISTRY

DOCTORAL THESIS

Development of novel methodologies for the identification of unknown compounds in the environment employing non-target screening and high-resolution mass spectrometry

NIKIFOROS ALYGIZAKIS

M.Sc. CHEMIST

ATHENS 2019

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ανάπτυξη καινοτόμων μεθοδολογιών ανίχνευσης αγνώστων ενώσεων στο περιβάλλον με μη στοχευμένη φασματομετρία μαζών υψηλής διακριτικής ικανότητας

ΝΙΚΗΦΟΡΟΣ ΑΛΥΓΙΖΑΚΗΣ

M.Sc. ΧΗΜΙΚΟΣ

ΑΘΗΝΑ 2019

DOCTORAL THESIS

Development of novel methodologies for the identification of unknown compounds in the environment employing non-target screening and high-resolution mass spectrometry

NIKIFOROS ALYGIZAKIS

Registration Number.: 150002

Supervising Professor:

Dr. Anastasios Economou, Professor, Chemistry Department, N.K.U.A.

Three – member consultative committee:

Dr. Jaroslav Slobodnik, Director Environmental Institute

Dr. Anastasios Economou, Professor, Chemistry Department, N.K.U.A.

Dr. Nikolaos Thomaidis, Professor, Chemistry Department, N.K.U.A.

Seven – member examination committee:

Dr. Jaroslav Slobodnik, Director Environmental Institute

Dr. Anastasios Economou, Professor, Chemistry Department, N.K.U.A.

Dr. Nikolaos Thomaidis, Professor, Chemistry Department, N.K.U.A.

Dr. Irene Panderi, Professor, Pharmacy Department, N.K.U.A

Dr. Manos Dassenakis, Professor, Chemistry Department, N.K.U.A.

Dr. Evagelos Gikas, Assistant Professor, Pharmacy Department, N.K.U.A

Dr. Yannis Dotsikas, Assistant Professor, Pharmacy Department, N.K.U.A

Defense Date: 15/04/2019

DOCTORAL THESIS

Ανάπτυξη καινοτόμων μεθοδολογιών ανίχνευσης αγνώστων ενώσεων στο περιβάλλον με μη στοχευμένη φασματομετρία μαζών υψηλής διακριτικής ικανότητας

ΝΙΚΗΦΟΡΟΣ ΑΛΥΓΙΖΑΚΗΣ

Αριθμός μητρώου: 150002

Επιβλέπων καθηγητής:

Δρ. Αναστάσιος Οικονόμου, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Τριμελής συμβουλευτική επιτροπή:

Δρ. Jaroslav Slobodnik, Διευθυντής του Environmental Institute

Δρ. Αναστάσιος Οικονόμου, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Δρ. Νικόλαος Θωμαΐδης, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Επταμελής εξεταστική επιτροπή:

Δρ. Jaroslav Slobodnik, Διευθυντής του Environmental Institute

Δρ. Αναστάσιος Οικονόμου, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Δρ. Νικόλαος Θωμαΐδης, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Δρ. Ειρήνη Παντερή, Καθηγήτρια, Τμήμα Φαρμακευτικής, Ε.Κ.Π.Α.

Δρ. Εμμανουηλ Δασενάκης, Καθηγητής, Τμήμα Χημείας, Ε.Κ.Π.Α.

Δρ. Ευάγγελος Γκίκας, Επίκουρος Καθηγητής, Τμήμα Φαρμακευτικής, Ε.Κ.Π.Α.

Δρ. Ιωάννης Ντότσικας, Επίκουρος Καθηγητής, Τμήμα Φαρμακευτικής, Ε.Κ.Π.Α.

Ημερομηνία κατάθεσης διατριβής: 15/04/2019

ABSTRACT

Wastewater treatment plants (WWTPs) are unable to remove many contaminants of emerging concern (CECs) efficiently, and therefore introduce them into the aquatic environment, where they form complex chemical mixtures containing typically thousands of individual substances. When analysed by liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS), these complex mixtures produce a high number of signals. To successfully translate these complex data into information required by environmental monitoring programmes, implementation of wide-scope target, suspect and non-target screening using powerful computational tools and related databases is required.

The objective of the thesis was to develop novel workflows employing state-of-the-art target, suspect and non-target screening tools and apply them on samples obtained from important European ecosystems such as the Danube River Basin (DRB) and the Black Sea (BS).

In this context, an introduction on CECs, workflows and techniques for their identification is presented in *Chapter 1,* which is followed by the scope of the thesis in *Chapter 2*. *Chapter 3* describes a non-target screening (NTS) workflow capable to prioritise compounds that exhibit large variation in their signal intensity over time (trend-analysis), which was used to detect events of direct disposal or sudden changes in the use of substances in WWTP of Athens. *Chapter 4* describes the establishment of a decentralised global emerging contaminant early-warning network to assess the spatial and temporal distribution using suspect screening. A platform to archive LC-HRMS data and apply wide-scope suspect screening of thousands of CECs, that incorporates all recent development in HRMS screening methods, is presented in *Chapter 5*. The platform was used to screen antibiotics and REACH chemicals in samples from BS (biota, sediment, seawater), various classes of CECs in wastewater from DRB (*Chapter 6*) and surfactants in wastewater samples collected within the national monitoring campaign in Germany (*Chapter 7*). Novel biomonitoring tools such as *in vitro* bioassays and analysis of antibiotic resistant genes (ARGs) supplemented NTS analyses of wastewater samples.

SUBJECT AREA: Analytical Chemistry

KEYWORDS: emerging contaminants, non-target screening, wide-scope target and suspect screening, Danube River Basin, Joint Black Sea Survey, German national monitoring campaign of wastewater effluents

ΠΕΡΙΛΗΨΗ

Τα κέντρα επεξεργασίας λυμάτων (ΚΕΛ) δεν απομακρύνουν αποτελεσματικά τους αναδυόμενους ρύπους. Έτσι, οι αναδυόμενοι ρύποι εισάγονται στο υδατικό περιβάλλον και σχηματίζουν πολύπλοκα μίγματα, τα οποία περιέχουν δεκάδες χιλιάδες χημικές ουσίες. Η ανάλυση αυτών των πολύπλοκων δειγμάτων με τεχνικές υψηλής απόδοσης, όπως υγροχρωματογραφία συζευγμένη με φασματομετρία μαζών υψηλής διακριτικής ικανότητας, παράγουν ένα μεγάλο αριθμό σημάτων. Για την επιτυχή μετάφραση των πολύπλοκων αυτών δεδομένων αλλά και για να επιτευχθεί ένα ολιστικό πρόγραμμα παρακολούθησης του περιβάλλοντος, απαιτείται ολόπλευρη χημική ανάλυση με εφαρμογή στοχευμένης σάρωσης, σάρωσης ύποπτων ενώσεων και μη στοχευμένης σάρωσης.

Ο σκοπός της Διατριβής είναι η ανάπτυξη εργαλείων ολόπλευρης χημικής ανάλυσης και η εφαρμογή τους σε σημαντικά Ευρωπαϊκά οικοσυστήματα όπως η λεκάνη απορροής του Δούναβη και η Μάυρη Θάλασσα.

Στα πλαίσια του σκοπού αυτού, στο *Κεφάλαιο 1* εισάγονται οι αναδυόμενοι ρύποι προτεραιότητας και οι τεχνικές ταυτοποίησης τους, ακολουθούμενοι από λεπτομερή περιγραφή των στόχων της Διατριβής στο *Κεφάλαιο 2*. Στο *Κεφάλαιο 3* περιγράφεται μια πορεία μη στοχευμένης σάρωσης για την προτεραιοποίηση ουσιών που παρουσιάζουν απότομη μεταβολή στην συγκέντρωση ως προς το χρόνο. Η πορεία αυτή χρησιμοποιήθηκε για την εύρεση φαινομένων απευθείας ρίψης ουσιών στο δίκτυο του ΚΕΛ της Αθήνας. Το *Κεφάλαιο 4* περιγράφει την ίδρυση ενός παγκόσμιου δικτύου έγκαιρης προειδοποίησης για την αξιολόγηση της χωρικής και χρονικής κατανομής νέων αναδυόμενων ρύπων. Εξέλιξη αυτού του δικτύου είναι ένα λογισμικό αποθήκευσης δεδομένων LC-HRMS με δυνατότητα εφαρμογής ευρείας αυτόματης σάρωσης ύποπτων ενώσεων, που ενσωματώνει όλα τα εργαλεία που χρησιμοποιούνται στις HRMS μεθόδους (*Κεφάλαιο 5*). Το λογισμικό χρησιμοποιήθηκε για την ανίχνευση αναδυόμενων ρύπων και χημικών ουσιών της βάσης REACH (i) σε δείγματα από τη Μαύρη θάλασσα (ζωντανοί όργανισμοί, ιζήματα και θαλάσσιο νερό), (ii) σε εξερχόμενα λύματα από τη λεκάνη απορροής του Δούναβη (*Κεφάλαιο 6*) και (iii) σε εξερχόμενα λύματα που συλλέχθηκαν από τη Γερμανία (*Κεφάλαιο 7*). Καινοτόμα εργαλεία βιοπαρακολούθησης όπως βιοδοκιμασίες και ανάλυση γονιδίων ανθεκτικών στα αντιβιοτικά συμπλήρωνουν τα αποτελέσματα της ανάλυσης αναδυόμενων ρύπων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Αναλυτική Χημεία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: αναδυόμενοι ρύποι προτεραιότητας, μη στοχευμένη ανάλυση, ευρεία σάρωση αναδυόμενων ρύπων, λεκάνη απορροής Δούναβη, δειγματοληψία Μαύρης Θάλασσας, εθνική δειγματοληψία Γερμανίας

To my family

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Dr. Jaroslav Slobodnik, who selected me from among hundreds of candidates to participate in the most prestigious European PhD programme under Marie Skłodowska-Curie PhD Fellowship. I would like to thank him for accepting me as a member of his team and for providing me with the opportunity to increase my knowledge and greatly expanding my scientific network. I would also like to acknowledge Prof. Nikolaos Thomaidis for his continuous support throughout my PhD and for helping me to develop my personality. Both, Dr. Jaroslav Slobodnik and Prof. Nikolaos Thomaidis are thanked for their excellent scientific guidance and for their trust in me. Additionally, I would like to thank Prof. Anastasios Economou for his help and encouragement throughout my PhD studies. Many thanks belong also to Dr. Irene Panderi, Dr. Manos Dassenakis, Dr. Evagelos Gikas, Dr. Yannis Dotsikas for agreeing to be in the examination committee and for their comments on my thesis.

I would also like to express gratitude to the leaders of laboratories that I had an opportunity to visit and work in during the secondments within the ANSWER project: Dr. Robert Mistrik (HighChem), Prof. Despo Fatta-Kassinos (UCY), Dr. Luc Hornstra (KWR), Dr. Peter Behnisch and Dr. Harrie Besselink (BDS). I would like to acknowledge all the support that I have received from Environmental Institute colleagues Dr. Peter Oswald, Martina Oswaldova, Michaela Mazanova, Peter Mazan and Dajana Vassova. My gratitude goes also to all my colleagues and friends in the team of University of Athens (TrAMS group) without whose help I would not make it for sure.

I would like to thank to all co-authors in our common papers for their exceptional scientific work and perfect collaboration. Special thanks go to the scientific community of the NORMAN network, especially Valeria Dulio, Juliane Hollender, Pablo Gago-Ferrero, Peter von der Ohe, Emma Schymanski and Kevin Thomas among many others. I would like to thank them for the collaborative efforts, their patience and persistency, and express the wish to stay together in the years to come, so that we can deal together with the up-coming scientific challenges.

Finally, I would like to thank my parents, my grandparents, and my wife for their patience and for their encouragement.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The experimental part of the thesis was conceived and performed in the laboratory of Environmental Institute (Slovakia) and in the Laboratory of Analytical Chemistry, Department of Chemistry of the National and Kapodistrian University of Athens (Greece). Part of the experiments were conducted in the laboratories of BDS and KWR (the Netherlands), whereas major data processing work took place at HighChem (Slovakia).

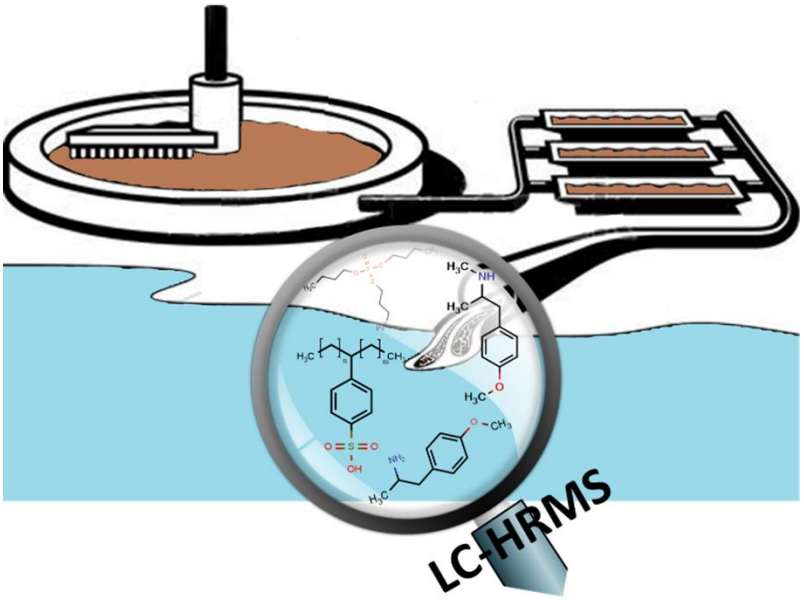
The thesis is accompanied by extended electronic supplementary information of 165 pages, consisting of 39 Figures and 29 Tables (available at http://users.uoa.gr/~nalygizakis/PhD/).

Studies presented in Chapter 4, 5, 6 and 7 received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 675530 (http://www.answer-itn.eu) and from the NORMAN Association under its Joint Programme of Activities (http://www.norman-network.net). The study presented in Chapter 3 was implemented under the Operational Program "Education and Lifelong Learning" and funded by the European Union (European Social Fund) and Greek National Resources - ARISTEIA 624 (TREMEPOL project, http://tremepol.chem.uoa.gr/).

Samples presented in Chapter 3 were collected in cooperation with Athens Water Supply and Sewerage Company (EYDAP S.A.) and samples for the study presented in Chapter 5 were provided by EU/UNDP EMBLAS-II project ENPI/2013/313-169 (http://emblasproject.org/) from the Joint Black Sea Survey 2016. Samples for the study presented in Chapter 6 were collected in cooperation with International Commission for the Protection of the Danube River (ICPDR) and samples for the study in Chapter 7 were provided by the German Environment Agency (Umweltbundesamt, UBA) under the study Project No. 94113.

CHAPTER 1

CONTAMINANTS OF EMERGINGCONCERN AND SCREENING METHODS FOR THEIR IDENTIFICATION



1.1 Introduction

The contamination of the environment, the food chain and drinking water is considered as a serious public health problem, which has been created by the ever-expanding world population. Thousands of chemicals, including pharmaceuticals, industrial chemicals, pesticides biocides and personal care products among others, find their way to the terrestrial and underground aquatic environment. Contaminants of emerging concern (CECs) enter the aquatic ecosystems from both diffused and point sources such as agriculture or wastewater treatment plants (WWTPs), respectively **[1, 2]**. WWTPs are being considered as the main contamination source of the aquatic environment and their incapability to efficiently remove the ever-increasing multitude of arriving chemicals is the cause of increasing concern **[3, 4]**.

According to European Chemicals Agency (ECHA) there are more than 68,000 registered chemicals under REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulation, which means that these chemicals are produced in Europe in quantities more than 1 tonne/annum **[5]**. Many of these chemicals and their transformation products (TPs) are persistent, potentially toxic and bioaccumulative **[6]**. Their concentration levels range from few pg L^{-1} to few μg L^{-1}. Despite their low concentration (most of them occur in few ng L^{-1} in

surface waters), their continuous introduction to the environment is of concern, because they are permanently present (pseudo-persistent), which might lead to unexpected chronic effects of affected species **[4, 7]**. Moreover, the occurrence of specific classes of emerging contaminants may trigger unknown and unexpected effects. An example of such unexpected effect is the spread of Antibiotic Resistance (AR) because of the occurrence of antibiotics, creating antibacterial conditions under which bacteria develop mechanisms of resistance against antibiotic drugs **[8]**.

European Union (EU) legislative efforts aim to achieve good chemical and ecological status for all European surface waters. Current legislation focuses on a list of 45 priority substances included in the Water Framework Directive (WFD) **[9]**, supplemented by a set of 15 additional compounds included in the recently revised watch list **[10]**. Member States are required to conduct monitoring programmes and benchmark the concentration levels of the priority substances against the Environmental Quality Standards (EQS) **[11]**. Despite the long-term efforts of the EU, water resources and aquatic ecosystems are still not sufficiently protected against the multitude of potentially hazardous substances **[12]**.

1.2 Contaminants of emerging concern

The term “emerging pollutants” or “emerging contaminants” or “contaminants of emerging concern” can be defined as chemicals that have been detected in the environment and potentially cause detrimental effects in the aquatic life at environmentally relevant concentration levels. They are pollutants that are currently not included in routine monitoring programmes at EU level and may be candidates for future regulation depending on their (eco)toxicity, potential health effects, public perception, and frequency of occurrence in environmental media. The fate, behaviour and (eco)toxicological effects are not well understood. This definition includes traditional substances, which are often present in the environment, but whose occurrence and significance are only now being evaluated. The organisation dealing with emerging substances is NORMAN Association, which is a unique network of reference laboratories, research centres and related organisations **[13, 14]**.

Emerging substances comprises a diverse group of compounds. It includes compounds of many classes and categories. Few of the most important categories of CECs are therapeutic

drugs prescribed for human (e.g. antibiotics, antihypertensive drugs, antiviral drugs, anticancer drugs and others), therapeutic drugs prescribed for animals (veterinary drugs), new psychoactive substances, industrial chemicals (further categorised to plasticizers, surfactants, food additives, dyes and pigments and others), chemicals used in agriculture (e.g. pesticides, insecticides, fungicides, biocides) and drugs of abuse **[15]**.

The term "emerging pollutants" also includes the TPs that are formed under biotic and abiotic transformation processes. The biotransformation products include human, animal and microbial metabolites in engineered and natural systems. The abiotic TPs are the outcome of abiotic processes such as hydrolysis, photolytic and photocatalytic degradation in the natural environment as well as water treatment processes, like chlorination, ozonation and advanced oxidation processes. TPs may differ in their environmental behaviour and ecotoxicological profile from the parent compound, depending on the modification that has taken place (e.g. oxidation, hydroxylation, hydrolysis, conjugation, cleavage, dealkylation, methylation and demethylation). In general, TPs are less toxic and more polar than the parent compounds. However, in some cases, they may be more persistent or exhibit higher toxicity or be present at much higher concentrations **[6]**. REACH regulation requires the identification of major TPs and degradation products for the registration of the substances **[16]**.

1.3 Instrumental analysis

Both liquid chromatography (LC) and gas chromatography (GC) have been used for the identification of emerging contaminants depending on their polarity, volatility and thermal stability. Polar emerging contaminants are mostly found in matrices such as surface water, ground water and wastewater, whereas non-polar emerging contaminants are mostly found in biota and sediment samples. An overlap between the two methods exists and collaborative trials organised within non-target screening cross-working group (NTS CWG) activity of NORMAN network have demonstrated that methods are complementary and the use of both of them in parallel is crucial to get a more complete overview on the presence of emerging substances in environmental samples **[17]**.

1.3.1 Gas chromatography coupled to mass spectrometry (GC-MS)

Gas chromatography coupled to mass spectrometry (GC-MS) is an established technique employed for the analysis of volatile and thermostable emerging substances. It has a high number of applications, which is a result of the well-developed mass spectral libraries and of the high sensitivity of the GC-MS methods, especially when combined with derivatisation of the analytes and multiple reaction monitoring methods (MRM) **[18, 19]**. The two most important GC-MS libraries (NIST and Wiley) contain more than one million mass spectra **[20]**. Therefore, the most common data processing of full-scan GC-MS data is to search and compare the obtained electron impact (EI) mass spectrum of a substance detected in the sample with mass spectra in the libraries. Compounds most commonly analysed by GC-MS include polycyclic aromatic hydrocarbons (PAHs), pesticides, polychlorinated biphenyls (PCBs), and numerous endocrine disrupting chemicals **[19]**.

1.3.2 Liquid chromatography coupled to mass spectrometry (LC-MS)

Liquid chromatography coupled to mass spectrometry (LC-MS) gained popularity in the last two decades, because it provided a platform applicable to a wide-range of polar and semi-polar compounds. Moreover, it provided high specificity, broad dynamic range and low sensitivity (especially in cases of triple quadrupole instruments). Soft ionisation interface such as electrospray ionisation (ESI) or atmospheric pressure chemical ionization (APCI) are commonly used to ionise the analytes. These methods let the molecular ion almost intact before reaching the detector and produce little in-source fragmentation. The fragmentation of the analytes required for structural elucidation is most commonly produced by collision-induced dissociation (CID), which can be performed in a specialised collision cell or in the intermediate-pressure part of the mass spectrometer (so called in-source CID). The most frequently used instrumental acquisition method is MRM **[21]**. In target screening, recording two MRM transitions between the precursor ion and the two most abundant product ions for each target analyte is considered sufficient for the identification and quantification of the target analytes (four identification points as described in 2002/657/EC) **[22]**. Identification of the structure of unknown compounds based on its isotopic profile, MS/MS or MS^n

fragmentation (in case of ion trap instrument) and the mass of the molecular ion is challenging due to the nominal mass accuracy of the MS instruments **[21]**.

1.3.3 Liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS)

LC-MS gained wide acceptance because of its capability to perform highly-sensitive quantitative measurements for few hundreds emerging substances in very complex environmental samples. However, it offered limited chances of successful identification of unknown compounds. Chances of structure elucidation and identification of unknown compounds were increased by the introduction of liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS). This instrumental setup gave to the researchers the opportunity to identify unknown compounds and TPs with high confidence and comparable sensitivity as LC-MS instruments. Nowadays, among the most common HRMS instruments have hybrid mass analysers such as quadrupole time-of-flight (Q-TOF), quadrupole-Orbitrap (Q-FT) and linear ion trap-Orbitrap (IT-FT) **[23, 24]**.

In these hybrid instruments, the first mass analyser is low resolution and the second is high resolution. TOF when combined with a reflector becomes a high-resolution mass analyser with resolution directly related to the length of the flight path. Contemporary instruments have a flight path of several meters, which is multiplied due to the reflector. Since resolution is related to the duration of flight time, TOF provides the highest resolution for relatively high *m/z* ion masses. On the contrary, orbitrap mass analysers provide the highest resolution for low *m/z* ions. Orbitrap has higher mass resolution than TOF, but this advantage is partially compensated by the lower speed of data acquisition. Moreover, the ratio of mass-to-peak width at full width at half maximum (FWHM) in TOF instruments is relatively constant over the entire mass range in contrast with the orbitrap analysers. Despite the differences in the quality characteristics, both mass analysers offer adequate mass resolution (R>30,000) and precise mass measurements (<5 ppm) at speed compatible with typical LC conditions **[25]**.

The hybrid instrumental setup allows the low-resolution mass analyser to pre-filter and isolate a mass of interest, which provides the opportunity for two main data acquisition methods: data independent acquisition (DIA) and data-dependent acquisition method (DDA) (**Figure 1**).

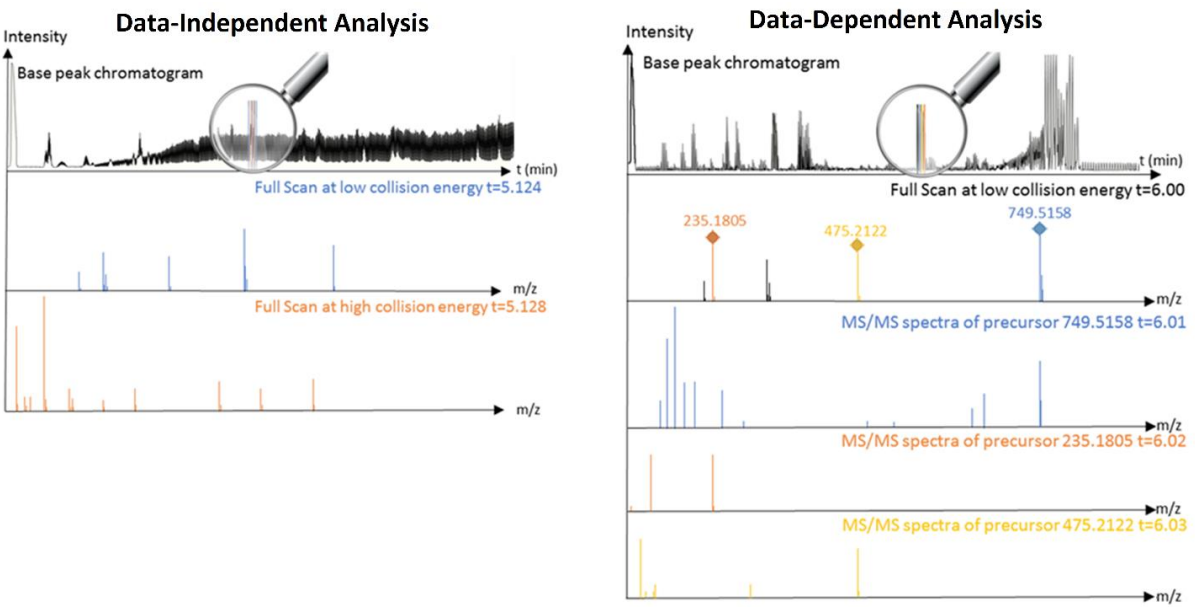


Figure 1. Visualisation of data-independent (DIA) and data-dependent (DDA) analysis methods

DIA, also termed as MS^E by Waters, bbCID by Bruker, ALLIONS by Agilent and MS/MS^{ALL} by Sciex, involves the sequential acquisition of accurate mass data at low collision energy and at high collision energy. In the low collision energy scan, almost no fragmentation takes place and thus molecular ions are recorded. In the high collision energy scan, all molecules are fragmented without any prior isolation of pre-selected ions. This, results in complex but rich in information MS/MS spectra, since they contain the fragmentation pattern of all co-eluting substances. DDA method records low collision energy full-scan, then, isolates and fragments pre-selected masses of interest (e.g. the three most-abundant ions as shown in **Figure 1**) at a given collision energy. MS/MS spectra from DDA method contain fragments only from the pre-selected ions. However, instruments are not fast enough to isolate and fragment all the observed molecular ions in a single chromatographic run. Multiple chromatographic runs are needed so that more DDA MS/MS scans are recorded, however, this requires sufficient quantity of the sample for analysis and multiple effort by the analyst for each analysed sample **[25, 26]**.

1.4 Identification approaches

1.4.1 Target screening

Target screening is the most common identification approach, in which reference standards are available. Thus, the retention time, and fragmentation pattern are known beforehand from the analysis of the reference standards. A compound is detected successfully by LC-HRMS through target screening in case there is a match of the retention time, the MS and the MS/MS fragmentation **[27, 28]**. In low-resolution LC-MS/MS instruments, target screening of compounds is achieved by the match of retention time and at least two MRM precursor-product transitions. In high-resolution LC-HRMS instruments, the limitation of the preselection of MRM transition has been eliminated, since LC-HRMS can operate in full-scan mode. This allows target screening of a theoretically unlimited number of compounds within a single chromatographic run. The limiting factor is the number of reference standards available in a laboratory. Wide-scope target screening of thousands of analytes is possible with DIA mode **[26, 29]**.

1.4.2 Suspect screening

Suspect screening refers to situations when no reference standards are available but there is prior structural information (mass spectra) on the suspects chemicals. In other words, suspect screening is the investigation of compounds with high chances of being present in the samples. Examples of suspected compounds are TPs of pharmaceuticals in influent wastewater and receiving waters and other compounds of a given class (e.g. suspect screening of surfactants) **[6, 28]**. Because the structures of suspected chemicals are known, a series of molecular properties can be directly derived: e.g. common adduct ions ($[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$ in positive and $[M-H]^-$ in negative ionization), *in-silico* predicted fragmentation pattern **[30-32]**, predicted retention time using Quantitative Structure (Chromatographic) Retention Relationship (QSRR) models **[33]** and predicted toxicity threshold using Quantitative Structure Toxicity Relationship (QSTR) models **[34]**. In case of the investigation of the occurrence of TPs, there are also tools available for predicting the structure of possible TPs (e.g. PathPred **[35]**, CATABOL **[36]** and enviPath **[37]**). The TP

prediction software contain rules of transformation and predict the TPs based on the parent molecular structure. In suspect screening, the detection of a chromatographic peak at plausible predicted retention time is not sufficient proof for identification of the related compound. The most important evidence is the plausibility of the fragmentation pattern – the 'fingerprint' of the compounds. The other supporting evidence for successful identification is a plausible isotopic profile, especially in cases in which the suspected substance contains atoms such as S, Cl and Br. Additional evidence that has been used to support identifications were metrics of wide-spread use of a suspected chemical (e.g. number of patents, number of data sources, number of references in chemical databases and publications) **[23, 38]**.

1.4.3 Non-target screening

In non-target screening no prior information and no reference standards are available. The first step in a non-target screening workflow is to apply an automated peak detection algorithm to detect chromatographic peaks in the three-dimensional (mass, retention time and intensity) LC-HRMS data. Since the peak picking will detect isotopic peaks and adducts, a componentisation algorithm to group peaks that belong to the same compound is applied, so that components are formed. Environmental samples are complex chemical mixtures containing thousands of individual substances. Their complete elucidation through the use of non-target screening is not feasible, since it would require extensive time and effort. Thus, it is clear that the selection of the peaks of interest (peak prioritization) is a key step in any investigation involving non-target analysis. Depending on the goals of the study, different prioritization strategies can be applied to the set of obtained components **[24, 39]**. So far, most of the prioritization strategies followed intensity-based criteria in combination with the prioritisation of substances with a distinctive isotopic pattern (e.g. halogenated compounds) **[40],** mass defect to focus identification efforts on molecular formulas outside the matrix domain in complex sediment samples **[41]** e.g. perfluoro-alkyl ether carboxylic acids and sulfonic acids in natural **[42]** and effect-directed analysis (EDA) **[43, 44].** Prioritised components were investigated using a non-target screening identification workflow.

The first step in the non-target screening identification workflow is the assignment of a molecular formula based on the exact mass, the isotopic fit and the exact mass of the fragments. There are software solutions offered by the MS manufacturers (e.g. Bruker SmartFormula **[45]**) and open-source software (e.g. Molgen MS/MS **[46]**), which are used to assign a molecular formula to a precursor mass. If the molecular formula is unequivocal, a chemical database (most commonly ChemSpider **[47]** and PubChem **[48]**) can be searched to retrieve candidate structures. Afterwards, either an open-source software such as MetFrag **[32]** or a commercial software such as MassFrontier **[31]** can be used to perform *in-silico* fragmentation of the candidate structures and match the predicted fragmentation with the observed fragmentation with the aim to rank the retrieved candidates. Search of the observed MS/MS spectrum in mass spectral libraries is a step not to be neglected, even though LC-HRMS mass spectral libraries contain a limited number of reference spectra **[49]**. Ranking of candidates can be influenced by a series of other available information (e.g. number of patents, references, presence of the candidate substances in different chemical libraries, etc.) **[23, 32]**.

1.5 Identification confidence levels in LC-HRMS

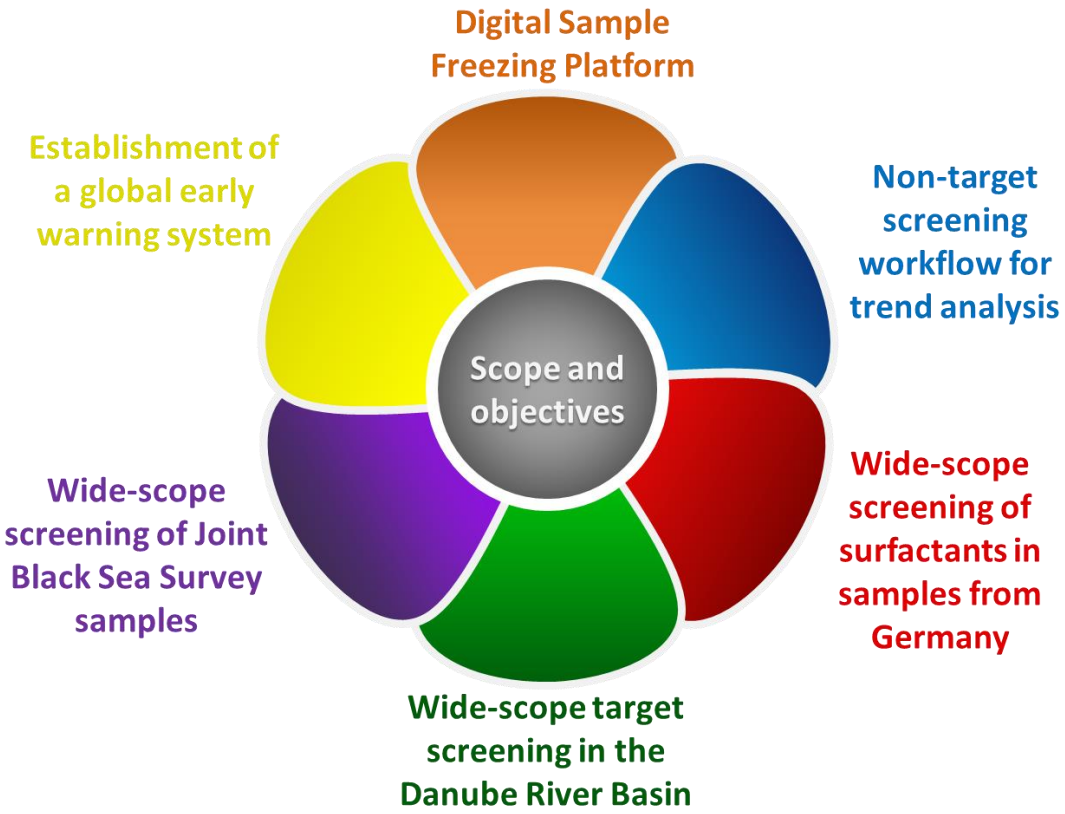
The level of confidence for the identification of the detected compounds should be clearly expressed based on the identification evidence. In environmental sciences, the most frequently used scheme consists of the following five identification levels **[27]**:

- **Level 1**: Confirmed structure is the case in which reference standard was purchased and the candidate structure is confirmed because of match of MS, MS/MS and retention time.
- **Level 2**: Probable structure refers to the proposal of an exact structure based on the following evidence:
    - **Level 2a**: Library spectrum matched with experimentally observed MS/MS spectrum
    - **Level 2b**: Diagnostic case, in which no other candidate fits in with the experimental data, but neither literature spectrum is available in mass spectral libraries nor reference standard is available

- **Level 3**: Tentative candidate(s) are proposed, because there is evidence, which is not enough to propose only a single exact structure.
- **Level 4**: Unequivocal molecular formula is proposed in cases in which there is limited experimental evidence to propose one or few structure(s), but there is enough evidence for assignment of an unequivocal formula
- **Level 5**: Mass of interest is proposed in case experimental information is not enough to conclude to a molecular formula

CHAPTER 2

SCOPE AND OBJECTIVES



2.1.1 The analytical problem

Thousands of CECs enter daily the environment due to anthropogenic activities, generating complex cocktails of chemicals, which may potentially harm the ecosystem and human health. Among the most important point sources of pollution are WWTPs, which are known to be unable to remove efficiently all the CECs. Moreover, TPs of the CECs, formed during the mechanical, biological and chemical processes at the WWTPs, find their way to the wastewater effluent. CECs and their TPs enter the water circle and have potential toxic effects on the aquatic ecosystem (e.g. algae, crustaceans and fish). Many of CECs and end up in humans either through the trophic chain or through the water itself.

To tackle this problem and protect its water bodies and citizens, EU water legislation (e.g. WFD **[9]**, EQS directive **[11]**), forces its Member States to run national monitoring programmes, with the aim to measure concentration levels of specific legacy pollutants, including the river basin specific pollutants and benchmark them against the established

EQSs. Despite the efforts of the policy makers, the legislation is limited on monitoring of a very few micropollutants, and overlooks the risk derived from the occurrence of thousands of unknown CECs.

Recent developments in advanced analytical instrumentation, especially in the field of HRMS have given the analytical scientists an opportunity to broaden their horizons. LC-HRMS has proven to be a powerful tool in hands of researchers to detect and reveal the identity of many unknown compounds in the environment. The high specificity and selectivity of hybrid mass spectrometers such as QTOF, QFT and ITFT instruments enabled them to deal with very complex matrices effectively.

The main difficulty of the advanced instrumentation is that the generated data is not possible to be fully interpreted and taken advantage of. Another obstacle is that there is no consensus among the vendor companies to use and follow an agreed data structure and format type. Instead, each HRMS vendor uses propriety data formats and have built their software based on them. Another important issue is that each vendor suggests to their users different HRMS acquisition methods for their analysis, which they brand under different names. All the above created a deficiency in the development of new data processing methods, able to give answers to the urgent environmental problems facing by the scientific community and the policy makers.

In the context of the presented thesis, two computational tools are presented, one of which has been developed in cooperation with the NORMAN network. A suggestion will be provided about the way how the aforementioned obstacles can be overcome. An emphasis will be given on how to organise holistic environmental monitoring programmes, which are prerequisite for dealing with the deterioration of the quality of the EU water resources due to the occurrence of CECs.

2.1.2 Research objectives and scope

The objective of this thesis was to develop novel methodologies for the investigation of the occurrence of CECs in the environment and apply them in European ecosystems. To achieve the objective, advanced analytical instrumentation and cutting-edge chemometric tools were developed and applied on the collected samples. The thesis is organised in five case studies, each one described in the following five chapters.

Chapter 3 describes a non-target screening workflow capable to prioritise compounds that exhibit large variation in their intensity over time (trend-analysis). The computational workflow is based on three open-source R packages (xcms, CAMERA and TIMECOURSE) and uses the statistical test Multivariate Empirical Bayes Approach (MEBA) to rank the components. The workflow was validated and was successfully applied in 8-day replicated composite flow-proportional influent wastewater from the WWTP of Athens. The top prioritised components were investigated using non-target screening identification workflow. Novel surfactant homologs were elucidated, four compounds were confirmed with reference standards and two were reported for the first time in the literature.

Chapter 4 describes the establishment of a decentralised global emerging contaminant early-warning system able to assess the spatial and temporal distribution of novel CECs using retrospective suspect screening. The study is the first pilot-scale joint activity, in which eight reference laboratories participated with available archived HRMS data. Data acquired from aqueous environmental samples collected in 14 countries from three continents were investigated. Laboratories compiled a suspect list of 156 analytes and investigated the occurrence of these emerging substances in their data. Widespread occurrence was proved for many compounds including industrial chemicals, TPs of pharmaceuticals and surfactants. All reported results were further examined through a quality control assessment and challenges derived from the quality control (QC) check were also discussed.

Chapter 5 presents a platform for archiving LC-HRMS data. The platform was used for the retrospective suspect screening of thousands of environmental pollutants with the ambition of becoming a European and possibly global standard. It was termed Digital Sample Freezing Platform (DSFP) and incorporates all the recent developments in the HRMS screening methods within the NORMAN Network. DSFP was used to screen for antibiotics and REACH

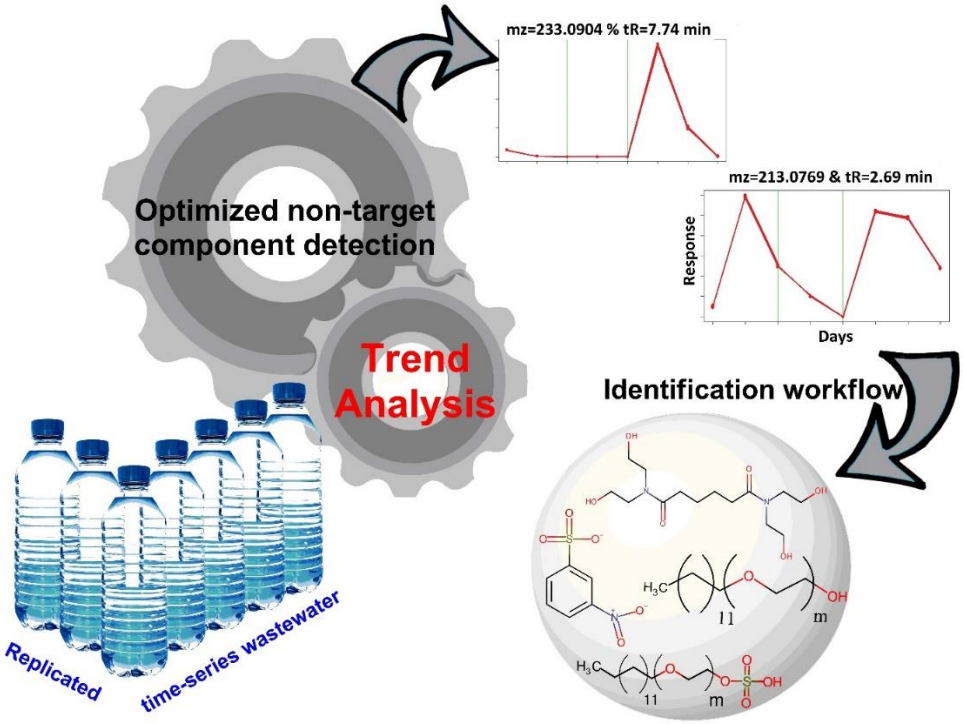
chemicals in seawater, sediment and biota samples collected during the Joint Black Sea Survey (JBSS).

Chapter 6 presents a study conducted in the Danube River Basin (DRB). Averaged 7-day composite effluent wastewater samples from 12 WWTPs in nine countries were collected. WWTPs' selection was based on countries' dominant technology and a number of served population with the aim to get a representative holistic view of the pollution status. Samples were chemically characterised using wide-scope target screening of 2248 emerging substances. Chemical analysis was supplemented by the application of a panel of *in vitro* CALUX® bioassays and analysis of 13 antibiotic resistant genes (ARGs) and one mobile genetic element (int1). All data gathered from these various analytical methods were stored in an-online interactive database. Additionally, risk assessment scheme was applied to prioritise hazardous substances and evaluate the signals from bioassays. A putative action plan based on the exceedance of effect-based trigger values (EBTs) was also proposed.

Chapter 7 describes the investigation of surfactants in wastewater effluents collected during the national monitoring campaign in Germany. Samples were analysed using target screening for linear alkylbenzene sulfonates (LAS), alkyl ethoxysulfates (AES) and using suspect screening through DSFP for 1,564 suspected surfactants and their TPs. Concentration levels of LAS were up to 14.4 µg/L, whereas AES occurred in lower concentration (up to 0.6 µg/L). Many LAS by-products and TPs such as di-alkyl tetralin sulfonates (DATSs), sulfophenyl alkyl carboxylic acids (SPACs) and sulfo-tetralin alkyl carboxylic acids (STACs) were detected and maximum semiquantified concentration levels reached 19 µg/L, 17 µg/L and 5.3 µg/L, respectively. Finally, suspect screening revealed the longest homologue series so far reported in the literature, consisting of 41 polyethylenoglycols (PEGs). Cumulatively, the concentration of surfactants in effluent wastewater reached up to 82 µg/L in a single sample.

CHAPTER 3

UNTARGETED TIME-PATTERN ANALYSIS OF LC-HRMS DATA TO DETECT SPILLS AND COMPOUNDS WITH HIGH FLUCTUATION IN INFLUENT WASTEWATER



Highlights

- Novel prioritisation capable of detecting compounds with high fluctuation over time
- Application to LC-HRMS data of daily influent wastewater samples
- 30% of the prioritised compounds were tentatively identified
- Two compounds were reported in wastewater for the first time
- Four novel surfactant series were tentatively identified

This case study has been published in Journal of Hazardous Materials, 5th January 2019, Volume 361, Pages 19-29 (10.1016/j.jhazmat.2018.08.073).

3.1 Introduction

Advances in high resolution mass spectrometry coupled to liquid chromatography (LC-HRMS) offer to environmental analytical chemists the opportunity to identify a continuously increasingly number of trace organic pollutants, even in highly complex environmental samples **[50]**. Target screening is insufficient to assess the quality of environmental waters as only a small portion of organic contaminants can be captured, while other relevant and potentially harmful substances cannot be detected **[28, 51, 52]**. Although still most investigations focus on target screening (where reference standards are available), there is an increasing number of studies dealing with both suspect screening (prior structural information of the suspects available, but no reference standards are available) and non-target screening (no prior information and no reference standards are available).

However, environmental samples are complex chemical mixtures containing tens of thousands of individual substances that produce a high number of peaks in LC-HRMS analysis. Their complete elucidation through the use of non-target strategies is not feasible, since it would require extensive time and effort. Thus, it is clear that the selection of the peaks of interest (*peak prioritisation*) is a key step in any investigation involving non-target analysis. Depending on the goals of the study, different prioritisation strategies should be applied to the set of obtained chromatographic peaks **[24]**.

So far, most of the prioritisation strategies followed intensity-based criteria in combination with the prioritisation of substances with a distinctive isotopic pattern (e.g. halogenated compounds) **[40, 52-54]**, as these can be considered as relevant substances with reasonable identification chances. Other approaches used mass defect to focus identification efforts on molecular formulas outside the matrix domain in complex samples **[41, 55].** It has also been proved useful when the objective is to find molecules with specific characteristics. An example was the detection of perfluoro-alkyl ether carboxylic acids and sulfonic acids in natural waters due to the negative mass defect of the multiple fluorine and oxygen atoms **[42].** Few studies conducted peak prioritisation prior non-target analysis based on effect-directed analysis (EDA), a useful tool for identifying predominant toxicants in complex environmental mixtures combining effect testing and fractionation **[43, 44, 56]**. Other strategies include time series prioritisation (prioritizing features whose intensities varied substantially over the time course of a sampling campaign in one sampling site) **[57, 58]**, are based on spatial variation **[59]** or

use metabolic logic combined with multivariate statistics in order to find unknown metabolites of certain substances **[60]**. In the field of trend analysis, Schlüsener et al. **[57]**, used vendor software from SCIEX (MarkerView) to analyse long-time series LC-HRMS data coming from a sampling station of Rhine river which was affected by effluent wastewater. Afterwards, they used open-source scripts to visualise the patterns and to perform autocorrelation to search and prioritise the features with high periodic variations**.** Plassmann et al. used trend analysis to detect continuously increasing peak intensities and filter out peak signals from naturally-occurring substances in whole blood samples **[58].** Moreover, trend analysis has been used for assessing the quality of the chromatographic stability in LC-HRMS data using von Neumann trend test [61, 62].

The main objective of the present study was the development of an automated prioritisation workflow based on open-source tools that is capable of detecting automatically compounds that exhibit large variation in their intensity over time (trend-analysis). This new prioritisation approach was realized by combining the different open-source R packages xcms, CAMERA and TIMECOURSE as well as the statistical test *Multivariate Empirical Bayes Approach* (MEBA) **[63].** The statistically obtained Hotelling T2 coefficient was used as an indicator of large intensity variations to rank the compounds. MEBA seemed to be the most suitable trend test for the generated dataset, because (i) it assesses longitudinal developmental time-series, (ii) it considers the repeatability of the replicates and (iii) it is not affected by progressive variations since data is not examined sequentially. Moreover, (iv) it accounts cumulatively for large variations among the different time points and (v) it is not affected by seasonality.

The developed workflow was applied to the evaluation of influent wastewater samples in order to detect events of direct disposal (e.g. due to illegal discharges) or sudden changes in the use of any substance. Replicated time-series of 24-hour flow proportional composite influent wastewater samples were taken during 8 consecutive days from a large wastewater treatment plant (WWTP) in Athens, Greece, which receives both urban and industrial wastewater. The compounds were ranked according to the developed procedure and elucidation efforts focused on the top-prioritised ones through the application of non-target identification strategies previously developed **[53]**.

3.2 Materials and methods

3.2.1 Chemicals and reagents

All solvents used in the present work were UPLC-MS grade. Acetonitrile (ACN) and methanol (MeOH) were purchased from Merck (Darmstadt, Germany), whereas 2-propanol of LC-MS grade was obtained from Fisher Scientific (Geel, Belgium). Distilled water was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA). Sodium hydroxide monohydrate (NaOH) for trace analysis ≥99.9995% and formic acid 99% were purchased from Fluka (Buchs, Switzerland). Details on the used chemicals and reagents for sample preparation and standard compounds purchased for confirmation purposes are provided in the **Supporting Information (SI) at section S3.1.**

3.2.2 Sampling and storage

24-hour composite flow proportional influent wastewater samples were collected from the WWTP of Athens (Greece) during 8 consecutive days in March 2015. The location of the WWTP of Athens can be found on **section S3.2 (SI).** The WWTP is designed with primary sedimentation, activated sludge process with biological nitrogen and phosphorus removal and secondary sedimentation. The residential population connected to the WWTP based on official census, excluding commuters, is 3,700,000 and the number of people estimated based on the number of house connections is 4,562,500. The WWTP is designed to serve a population equivalent of 5,200,000 and thus is by far the largest in Greece and one of the largest in the world. The estimated sewage flow for the collected samples was 720,000 m^3 day^{-1}.

Raw influent wastewater was collected in pre-cleaned high-density polyethylene (HDPE) bottles. The samples were filtered with glass fiber filters (pore size 0.7 μm) immediately after arrival at the laboratory. They were stored in the dark at 4 °C until analysis, which happened directly after the end of the sampling campaign.

3.2.3 Sample preparation and instrumental analysis

Sample extraction was carried out using a slightly modified protocol developed by Kern et al. **[64]**. In-house four sorbent SPE cartridges (200 mg Strata-X, 150 mg Isolute ENV+, 100 mg

Strata-X-AW and 100 mg Strata-X-CW) were preconditioned with 6 mL with MeOH and 6 mL water. Cartridges were loaded with aliquots of 100 mL (preadjusted to pH 6.5), were dried under vacuum for 1 hour and were eluted with 4 mL of 50:50 MeOH:ethyl acetate containing 2% of ammonia, followed by 2 mL of 50:50 MeOH:ethyl acetate containing 1.7% of formic acid. Extracts were evaporated under a gentle nitrogen stream to a volume of 100 µL, reconstituted to 0.5 mL with a final proportion of 50:50 MeOH:water and filtered through a 0.2 µm RC syringe filter (Phenomenex, USA) .

Analyses were carried out using an UHPLC-QTOF-MS system, equipped with a UHPLC apparatus (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Germany), consisting of a solvent rack degasser, auto-sampler, a binary pump with solvent selection valve and a column oven coupled to the QTOF-MS/MS analyzer (Maxis Impact, Bruker Daltonics, Bremen, Germany). An Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 µm) from Thermo Fisher Scientific (Dreieich, Germany), preceded by an ACQUITY UPLC BEH C18 1.7 µm, VanGuard Pre-Column from Waters (Dublin, Ireland), and thermostated at 30 °C, was used for separation.

All the samples were first analysed in full scan mode. The QTOF-MS system was operating in broadband collision-induced dissociation (bbCID, data-independent) acquisition mode and recorded spectra over the range of *m/z* 50−1000 with a scan rate of 2 Hz. This mode provides MS and MS/MS spectra at the same time working at two different collision energies (4 and 25 eV). A second data-dependent MS/MS acquisition was conducted using a preselected inclusion mass list containing the exact masses of the precursor ion of selected compounds. The collision energy applied was set to predefined values, according to the mass and the charge state of every ion. Detailed information on the UPLC-MS/MS performance is provided in **section S3.3 (SI)**.

3.2.4 Computational workflow

Raw files acquired from the LC-HRMS analysis were converted to mzML file format by using Proteowizard software **[65]** with the following conversion parameters: *Peak Picking,* true 1-; *MsLevel,* 1-1 and *Threshold peak filter,* absolute 300-most intense. The computational workflow and the prioritisation methodology here-in proposed is based on functions available in three R-packages. In brief, functions for peak detection, matching peaks across the samples

and OBI-Warp retention time alignment are included in the XCMS R package, while functions for componentization based on retention time and peak shape and functions for annotation of adducts and isotopic peaks are included in the CAMERA R-package. TIMECOURSE package was used for prioritisation using the *one sample multivariate empirical Bayes statistic* developed by Tai and Speed **[63]**. A step-wise illustration of the computational workflow can be found at **Figure 2.**

Sample feature detection was the first step and it was carried out using the function xcmsSet() with optimized parameters for QTOF MS data (CentWave parameters can be found in **Table 1**). After that, features representing the same analyte across samples were placed into groups using the group() function. Retention time alignment was performed using retcor() function (based on the Kernel density estimator **[66]**). Since there were feature groups with missing features from some of the samples (e.g., because an analyte is not present in a particular sample), these missing features were filled with a low intensity value with fillPeaks() function **[67]**. This is important in order to avoid errors due to missing values of non-detected peaks in some samples, when performing statistical analysis. Then, features were clustered according to retention time (using groupFWHM() function) and further according to the peak shape correlation coefficient (using groupCorr () function). For this purpose xcmsSet objects were converted to CAMERA objects by using xsAnnotate() function. Finally, isotopic peaks and adducts were annotated using the functions findIsotopes() and findAdducts(), respectively **[68]**. Peaks detected in the blank samples (with an intensity ratio below one order of magnitude) were removed. Target compounds were excluded based on accurate mass (±mass accuracy window of 3 mDa) and retention time (±retention time window of 0.50 min). Discussion of target screening results is out of the scope of the present manuscript. All remaining components were normalized by log2 transformation. After that, the statistical test (Multivariate Empirical Bayes Approach **[63]**) was applied and compounds were ranked based on the Hotelling T2 coefficient, which can be used as an indicator of large concentration variations among daily composite samples.

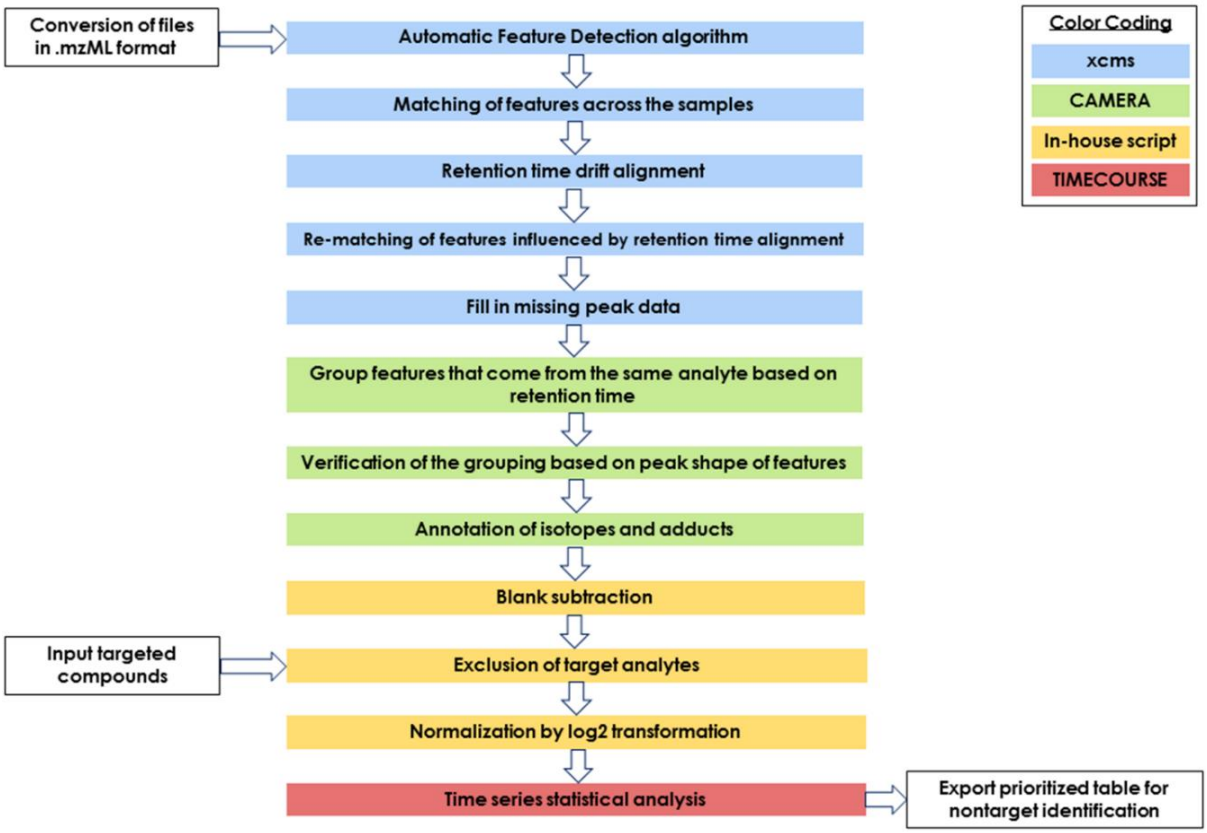


Figure 2. Compiled and optimized workflow for detecting compounds with a characteristic intensity fluctuation over time.

3.2.5 Identification of unknown compounds

Identification of top prioritised components was based on the non-target approach established by Gago-Ferrero et al **[53]**. Possible molecular formulas were assigned by applying thresholds of mass accuracy (≤2 mDa) and isotope pattern (mSigma≤50 **[69]**). If elucidation of the molecular formula was not unequivocal based on mass accuracy and isotope pattern, MS/MS was also considered using Molgen-MS/MS software **[46]**. Molgen-MS/MS was used with the parameter following settings: Elements - C, H, N, O, P, S (unless there was evidence of halogens), existence filter "exist", odd electron ions (oei), ppm = 5 and acc = 15 (MS and MS/MS accuracy settings in ppm). Once determined the molecular formula, candidates were obtained through the evaluation of the MS/MS spectra, including the use of in silico fragmentation platforms (Metfrag **[70]** via Metfusion **[71]**) and the MassBank library **[72]**. Commercial importance criteria was also used through the evaluation of the number of references and data sources in ChemSpider **[47]** and the number of patents in PubChem **[48]**. The chromatographic retention time plausibility of the candidates was evaluated, using an in-house QSRR retention time prediction model **[33]**.

In four cases, the identity of the unknowns was confirmed by purchasing the corresponding standard and comparison of the t_R and MS/MS spectrum. Spectral similarity values were calculated with the OrgMassSpecR package in R **[73, 74]**. Confirmation was considered successful only when t_R deviation was below 0.2 min and MS/MS spectrum similarity was higher than 70%. The level of confidence for the identification of the detected compounds was used according to Schymanski et al. **[27]**, where Level 1 corresponds to confirmed structures (reference standard is available), level 2 to probable structures, level 3 for tentative candidate(s), Level 4 to unequivocal molecular formulas, and level 5 to exact mass(es) of interest.

3.3 Results and discussion

3.3.1 Optimization of the computational workflow to obtain component lists

The computational workflow established in order to obtain the compound list consists of three basic steps: peak picking, matching peaks across the samples and chromatographic t_R alignment. Different input parameters in the aforementioned steps (e.g., mass accuracy or

peak width in centWave peak picking algorithm) may lead to different compound lists **[75, 76]**. Therefore, parameters were optimized by Box-Behnken fractional factorial design (IPO R-package) **[75]**. Optimized values for each parameter are summarized in **Table 1** and are discussed below.

IPO optimization is based on natural stable ^{13}C isotopic peaks. It calculates a peak picking score based on reliable peaks, meaning peaks for which their corresponding isotopes have been detected. This score combined with the total number of detected peaks and the number of low intensity peaks (isotopes may remain undetectable) is used as response variable. Peak picking parameters are tuned, so that the response variable is maximized following design of experiments method [75]. Optimum values for mass accuracy and peak width were almost the same in positive and negative ionisation mode (17.6 ppm, ~15 sec (*minimum peakwidth*) and 50 sec (*maximum peakwidth)*). The similarity in ESI(+) and ESI(-) mode was expected since the same separation method and instrument was used for analysis of the extracts in both polarities. The obtained mass accuracy threshold is lower than those used in most of the predefined methods in R-based online platforms (Scripps center for metabolomics (xcmsonline) **[77]**), where normally ~ 30 ppm is applied for QTOF data, and therefore decreasing the number of false positives. This example shows that optimizing input parameters prior to data treatment is important for proper dataset generation and therefore prioritisation. Other additional filters included such as *prefilter*, which is used in order to avoid peaks with very low intensity. When applying this filter, a given mass should be present at least in three consecutive scans with an intensity threshold (≥3000, ESI(+) and ≥1000, ESI(-)). Another filter was *scanrange*, which helps to avoid calibrant peaks by restricting peak picking to specific time intervals. In our case, calibrant substance was injected in the beginning of each chromatographic run using a 6-port valve and calibrant peaks appear for 12 consecutive scans, which were excluded by using the *scanrange* filter.

Table 1. Parameters used for the computational analysis.

Input Parameter	Positive ESI	Negative ESI
CentWave parameters		
ppm	17.6	17.6
Minimum peak width	14.34	15.5
Maximum peak width	50	50
prefilter	3, 3000	3, 1000
scanrange	20 until 1840	20 until 1840
fitgauss	TRUE	TRUE
integrate	TRUE	TRUE
Retention Time alignment based on OBI-Warp algorithm		
Distance function	cor_opt	cor_opt
gapInit	0.3	0.27
gapExtend	2.4	2.36
Grouping of features based on kernel density estimator		
bw	5	5
mzwid	0.032	0.0305
minfrac	0.6	0.6
minsamp	2	2
max	50	50

Only features existing in 3 out of the 5 replicates were kept by setting the parameter *minfrac* (minimum fraction of samples in a subgroup) to 0.6. After that, the kernel density estimator method was used for matching peaks across the samples (grouping together peaks representing the same analyte in different samples). In this regard, the parameters *bw* (bandwidth of kernels) and *mzwid* (width of overlapping *m/z* slices), which indicate time tolerance and mass accuracy, respectively, were optimized (**Table 1, part grouping of features based on kernel density estimator**).

The next step consisted of retention time alignment. It was performed by using the *ordered bijective interpolated warping (OBI-Warp)* algorithm **[66]**. Two penalty parameters, *gapInit* and *gapExtend*, which prevent the over-alignment of the chromatograms, were optimized. Optimization of grouping and retention time alignment takes place at the same time and is based on peaks appearing in all samples. Response variable is a linear combination of grouping response variable and retention time alignment response variable [75]. The obtained values were very similar to those obtained by Prince and Marcotte **[66] (Table 1)**. Moreover, it was observed that the maximum chromatographic drift during the analysis was 20 and 10 s (ESI(+) and ESI(-), respectively), showing the robustness of the chromatographic system.

Since ESI is a soft ionisation technique, several ion species can be observed for the same compounds (e.g., adducts or isotopes). In order to obtain the final compound list, the peaks belonging to the same compound were grouped. This was conducted using the CAMERA R-package **[68]**. This package can group the peaks based on retention time and peak shape and annotates isotopic and adduct peaks. Finally, to avoid prioritizing known substances, 207 and 32 target components in positive and negative ionisation respectively were excluded (target list of University of Athens consisted of 2248 compounds and is available at NORMAN Suspect list exchange http://www.norman-network.com/?q=node/236).

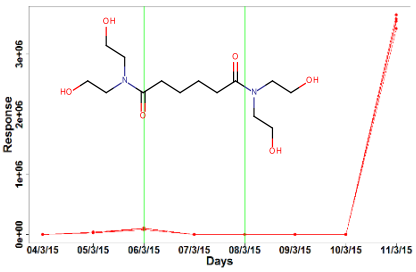
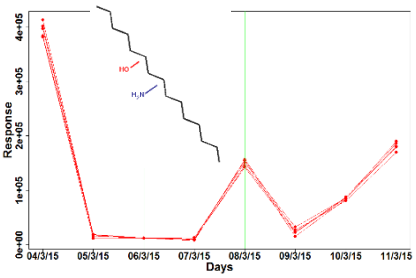
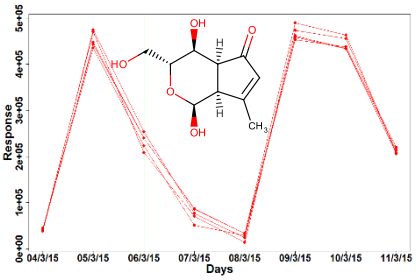
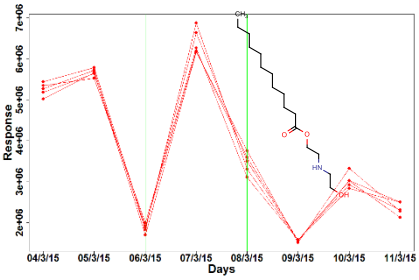
3.3.2 Prioritisation methodology

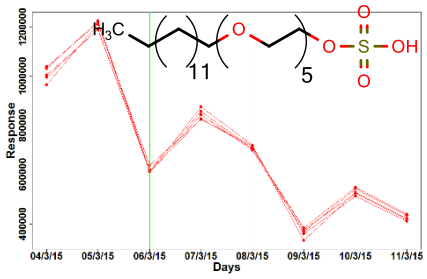
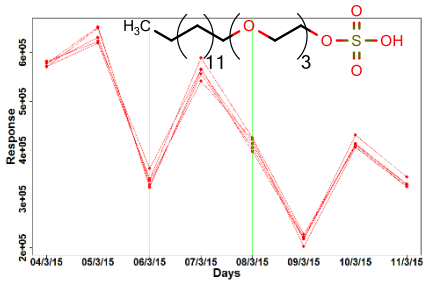
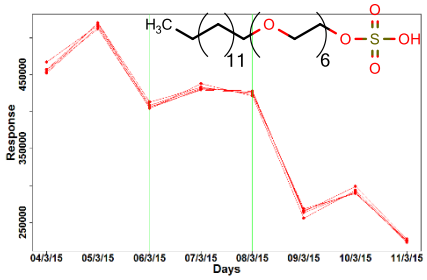
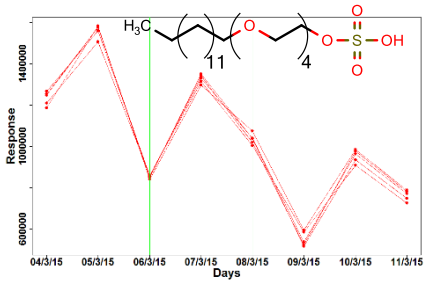
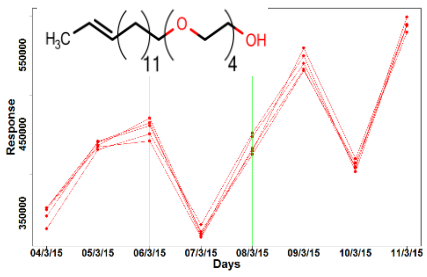
To find the compounds exhibiting high fluctuation among the daily samples, the *one-sample Multivariate Empirical Bayes Approach (MEBA)* statistical test was applied. This test is suitable for longitudinal replicated developmental time-course data. Originally, this statistical test was designed to solve the problem of ranking genes in microarray experiments **[63]**. MEBA has

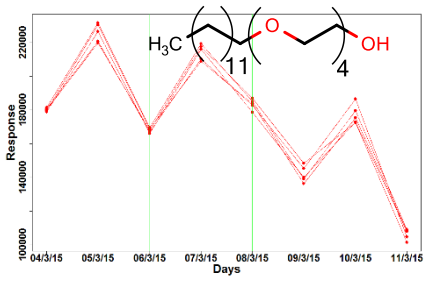
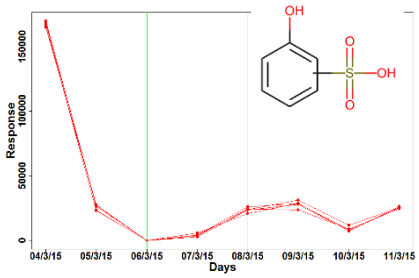
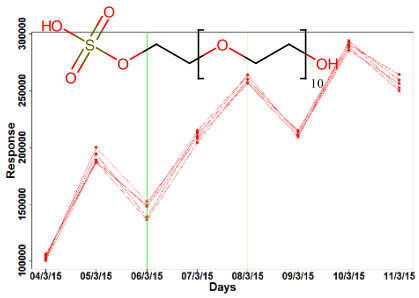
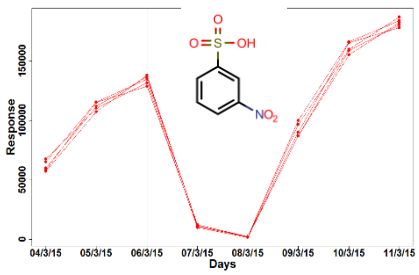
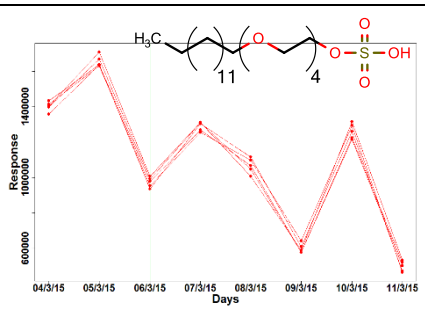
advantages compared to other F-statistic approaches (i.e. ANOVA) since it incorporates replicate variances and the correlations among responses of time-series samples from longitudinal data.

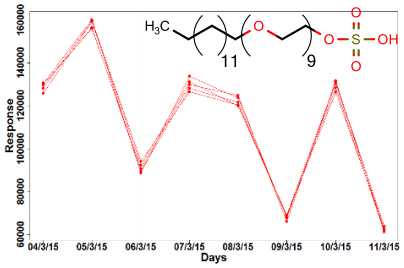
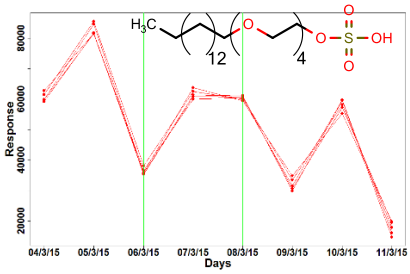
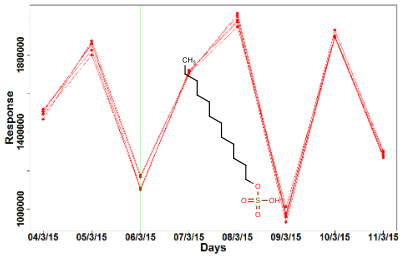
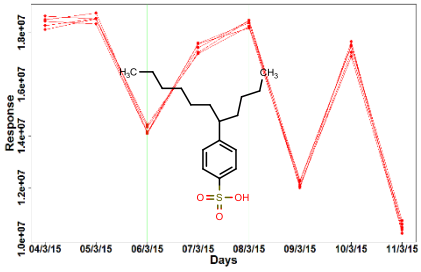
Intensity normalization is a mandatory step for statistical hypothesis testing. Therefore, as a first step Log2 transformation was performed in the dataset (compound list), since it is the most appropriate transformation for the applied statistical test **[63]**. Then, the statistical test was applied to every compound and a score (Hotelling T2) was assigned based on the peak area values observed in the time-series samples. This score is a positive number without an upper limit, which takes into account the repeatability of the intensity among replicates representing one time point and the magnitude of change of intensity between time points. A high value indicates high fluctuation among the time series samples. Compounds were ranked according to the score and the results for the first top 30 prioritised substances in each ionisation mode are summarized in **Table 2** and in **SI (section S3.4, Tables S3-4A and S3-4B**).

Table 2. Summary of the results for the prioritised and tentatively identified compounds in positive and negative ionisation mode.

Rank	m/z	Molecular formula (Name if available)	t_R (min)	Pred. t_R (min)	Level of confidence	Time trend
#P2	321.2033	$C_{14}H_{28}N_2O_6$ N,N,N',N'-Tetrakis(2-hydroxyethyl) hexanediamide)	2.70	2.80	1	
#P3	259.2822	$C_{16}H_{35}NO$	12.98	-	3	
#P9	215.0916	$C_{10}H_{14}O_5$ (1S,3R,4S,4aS,7aS)-1,4-dihydroxy-3-(hydroxymethyl)-7-methyl-3,4,4a,7a-tetrahydro-1H-cyclopenta[c]pyran-5-one)	3.40	3.34	3	
#P16	288.2539	$C_{16}H_{33}NO_3$ (2-[(2-Hydroxyethyl) amino] ethyl laurate)	11.74	9.84	3	

#P19	525.2544	$C_{22}H_{46}O_9S$	12.81	-	3	
#P20	437.1973	$C_{18}H_{38}O_7S$	12.66	-	3	
#P21	569.2756	$C_{24}H_{50}O_{10}S$	12.89	-	3	
#P22	481.2233	$C_{20}H_{42}O_8S$	12.62	-	3	
#P27	726.5726	$C_{38}H_{76}NO_{11}$	14.98	-	3	

#P29	363.3105	$C_{20}H_{42}O_5$ (3,6,9,12-Tetraoxatetracosan-1-ol)	13.51	13.05	1	
#N1	172.9914	$C_6H_5O_4S$	3.27	-	3	
#N3	581.2464	$C_{22}H_{46}O_{15}S$ (10 GES)	4.30	7.90*	2B	
#N16	201.9817	$C_6H_4NO_5S$ (3-nitrobenzenesulfonic acid)	3.08	3.17	1	
#N21	441.2546	$C_{20}H_{42}O_8S$ (3,6,9,12-Tetraoxatetracos-1-yl hydrogen sulfate)	12.46	10.11*	3	

#N22	661.3863	$C_{30}H_{62}O_{13}S$ (3,6,9,12,15,18,21,24,27-nonaoxanonatriacontyl hydrogen sulfate)	12.73	11.21	3	
#N24	455.2694	$C_{21}H_{44}O_8S$ (2-[2-[2-(2-tridecoxyethoxy) ethoxy] ethoxy] ethyl hydrogen sulfate)	12.98	10.20	3	
#N27	265.1457	$C_{12}H_{26}O_4S$ (lauryl sulfate)	10.94	10.05	1	
#N28	311.1684	$C_{17}H_{28}O_3S$ (C11-LAS)	12.00	12.10	3§ MassBank record: ETS00014	

§mix of isomers (spectra present in MassBank as a mix of isomers)

*Out of the domain of the retention time prediction mode

Through the evaluation of the graphs several compounds with a pollution spill trend could be observed. The graphics for the compounds with this behaviour are summarized in the **SI (Figure S3-5A)** and in **Figure 3** (selected cases).

Spill trend cases were compounds detected in specific samples (normally at high intensities), while remain undetectable in the other samples. This becomes obvious for the top-ranked compounds and especially for the cases **#P1, #P2** (**Figure 3**) as well as for the others depicted in **Figure S3-5B (SI)**, which exhibit extreme changes in intensities and were mainly found in one daily sample. Cases of pollution spills can also include compounds that can be detected in most of the samples at low intensity but the signal increase disproportionately in specific samples. The most obvious cases are **#P12** and **#P10**, where the signal increased more than 5 and 18 times, respectively, compared to the average intensity. Compounds belonging to this pollution spill category are of crucial environmental importance, since they can reach high concentration levels and become potentially toxic for the ecosystem. The detection and identification of these substances may allow the authorities to trace the pollution source and adopt appropriate measures.

Apart from the cases of pollution spills, also compounds with dropping signal intensities during specific days were determined through the application of the developed prioritisation methodology. Examples of this behaviour for the compounds **#N18** and **#N16** are depicted in **Figure 3B.** The signal decreased very significantly during the weekend period indicating an industrial origin.

Several of the prioritised compounds corresponded to substances exhibiting the same time pattern. In almost all cases of successful identifications these substances were identified as surfactants belonging to different homologue series. **Figure 3C** shows an example with three different surfactants sharing the same time pattern. More examples of this behaviour are depicted in **Figure S3-6B (SI)**. These compounds obviously share a common origin and even might coexist in products. However, in order to draw sound conclusions with other groups of substances more successful identifications would be required.

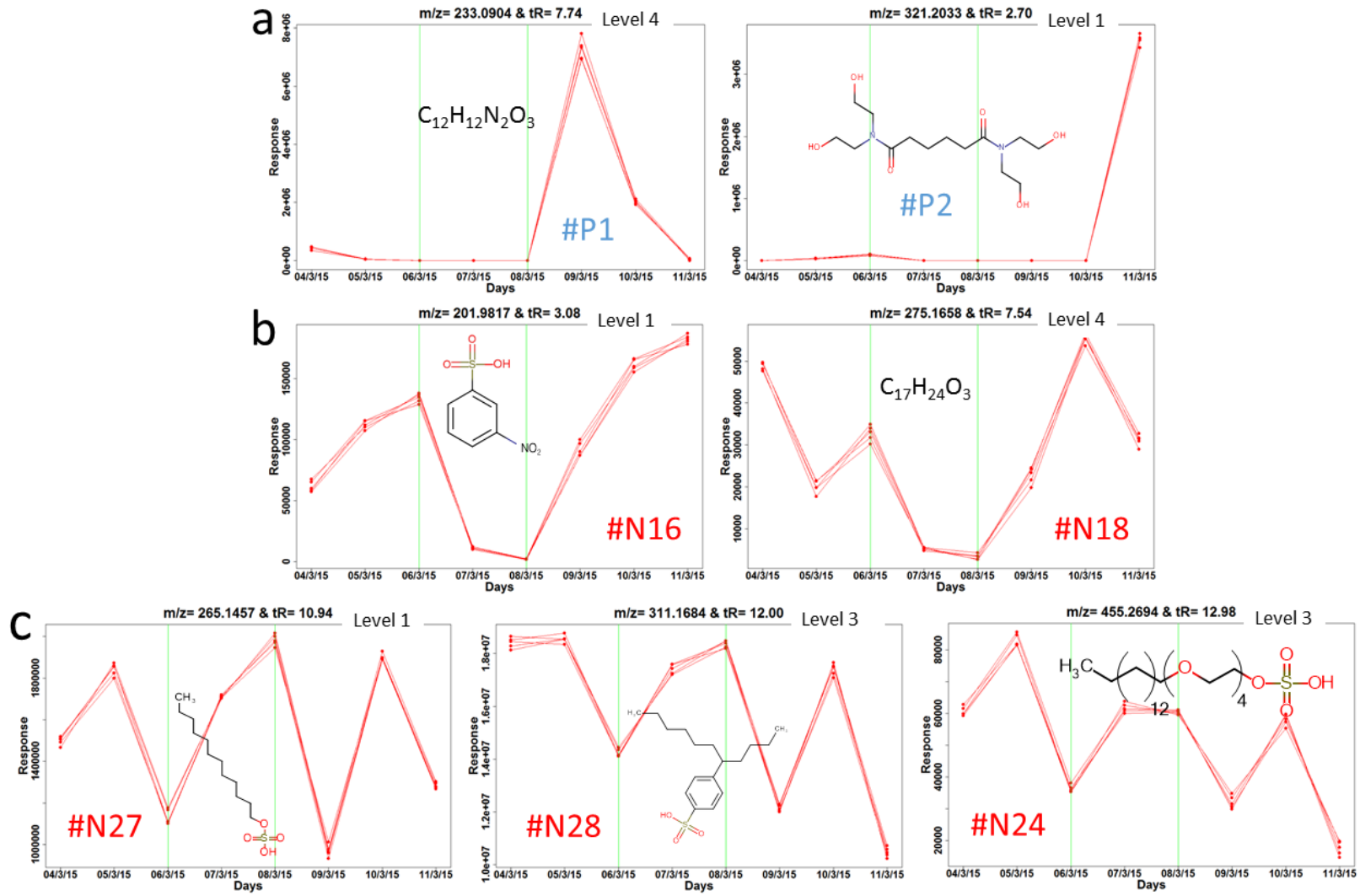


Figure 3. (A) Examples of pollution spills (events of direct disposal of chemicals into the sewage system); (B) Examples of compounds with dropping response during the weekend (The space between the green lines correspond to the weekend); (c) Surfactant compounds sharing a common time trend.

Evaluating the ranking, substances that remained undetectable in at least one sample (very low number or zero assigned by fill gaps function) also received a relatively high score, since the statistical approach is highly affected by this fact. This is the reason why some compounds (e.g. **#N15** or **#P7**) were prioritised even though their pattern in the rest of the time-points seems almost steady. Also, the score of a compound decreases when the repeatability within replicates of a sample time point is low, since MEBA takes into account all the replicates. This is the main reason why compounds with very similar trends are ranked slightly different (e.g. **#N9-#N12, SI**). Despite of the aforementioned disadvantages the prioritisation approach provided good results and it was proved capable of detecting pollution spills and compounds exhibiting high fluctuation over time.

3.3.3 Identification of top-ranked prioritised compounds

Identification efforts were focused on the first 30 prioritised components in each ionisation mode (60 potential compounds in total) and the results are summarized in **Table 2** (tentatively identified compounds) and **Table S3-4A**, **S3-4B (SI)** for prioritised but not tentatively identified compounds). In ESI (+), two substances, 3,6,9,12-tetraoxatetracosan-1-ol and N,N,N`,N`-tetrakis(2-hydroxyethyl)hexanediamide, were confirmed with the corresponding commercial standard, reaching the confidence level 1. Eight additional substances were tentatively identified; all of them as tentative candidates (level 3). For eight compounds it was not possible to go beyond the determination of the unequivocal molecular formula (level 4) and the remaining twelve compounds remained as exact mass of interest (level 5). In ESI (-), two substances, 3-nitrobenzenesulfonic acid and lauryl sulfate, were confirmed, one reached confidence level 2 and five compounds reached level 3. For twelve additional compounds an unequivocal molecular formula was assigned (level 4) and ten peaks remained at level 5.

An interesting case was the identification of the compound N,N,N`,N`-tetrakis(2-hydroxyethyl)hexanediamide (CAS: 6334-25-4) (case **#P2** in **Table 2**). A peak corresponding to *m/z* 321.2033 (t_R 2.70 min) was prioritised and the unequivocal molecular formula $C_{14}H_{28}N_2O_6$ was assigned based on the mass accuracy, the isotope

pattern and the annotation of the fragments. There were 38 compounds with this formula in the ChemSpider database. The MS/MS spectra indicated a neutral loss of 105.078 (corresponding to $C_4H_{11}NO_2$) and the loss of a H_2O molecule. The structure corresponding to the confirmed compound received the highest MetFrag score and was within the top 4 MetFusion candidates. Moreover, this compound was the one with the highest commercial importance (38 data sources, 41 references and 7 patents in ChemSpider and PubChem, respectively) in comparison with the other candidates. In addition, the confirmed compound received the closest predicted retention time, indicating that models for prediction of chromatographic behaviour can be useful for helping in revealing the identity of unknown compounds. Finally, the identity of the substance was confirmed with a commercial standard. This substance was present in 3 out of the 8 evaluated days, two of them at an almost negligible intensity (~ 3 until 9×10^4) and at very high intensity on the other day (Wednesday 11th March 2015 (3.6×10^7). Therefore, this is a characteristic example of a pollution spill-trend. This chemical is mainly used in the fabrication of adhesives, where it is added in order to enhance their performance by acting as cross-linker **[78]**. An intensive use of this substance during the specific day of 11th March 2015 by some adhesive industry with resulting high concentrations in the discharged wastewater or an event of direct disposal of this chemical into the sewage system are plausible hypothesis to explain the observed behaviour. Another interesting example of a compound with a pollution spill trend can be found in the case **#N1** (**Table 2** and **Figure S3-6B (SI)**). This compound was tentatively identified as hydroxybenzenesulfonic acid (level 3). On a specific day (Wednesday, 4th of March 2015), it was determined at a concentration 5 times higher than the average of the remaining days of the sampling campaign.

A compound showing lower concentration levels during the weekend was 3-nitrobenzenesulfonic acid (CAS 98-47-5, case **#N16**, **Table 2**), which was confirmed with a standard, reaching level 1. This compound is used in electrical/electronics, photographic, and textile processing industries **[79]**. The behaviour of this compound can be explained due to the fact that these industries do not operate (at least at the same level) during weekends leading to decreasing concentrations. Another substance which also showed lower levels during the weekend days was tentatively

identified (level 3) as the glucuronated derivative of 3-methylcyclopent-2-enone (CAS: 251914-61-1) (case **#P9**). This chemical is used in the food industry as a colour additive **[80]**. For other compounds following exactly the same trend (summarized in **Figure S3-5C (SI)**), it was not possible to go beyond level 4, mostly due to the high number of potential candidates.

Several surfactants belonging to the homologous series $CH_3(CH_2)_{11}(CH_2CH_2O)_XSO_4$ (X=1…12), which have not been previously reported in wastewater, were detected and identified as it is shown in **Figure 4**. These compounds corresponded to the cases **#P20** m/z 437.1973 (t_R: 12.66 min), **#P22** m/z 481.2233 (t_R: 12.62 min), **#P19** m/z 525.2544 (t_R: 12.81 min) and **#P21** m/z 569.2756 (t_R: 12.89 min) (**Table S3-5B (SI)**). In both ionisation modes, consistent peak shapes and constant increase of t_R were observed when increasing the chain length. The proposed structures can explain all the fragments obtained in the ESI(+)-QTOFMS (**Figure 4**). All the spectra corresponding to the homologous series were very similar, showing in all cases characteristic fragments at *m/z* 45.0334, 89.0597, 133.0859 and 177.1121, corresponding to the group $(CH_2CH_2O)_x$ (x=1-4). The protonated adduct could not be detected in ESI(+)-QTOFMS. However, the adducts $[M+NH_4]^+$ and $[M+K]^+$ showed high intensity, in agreement with other studies dealing with identification of surfactants **[53, 81]**. Substances with the same molecular formulas and time trend were also detected in ESI(-)-QTOFMS. The presence in the MS/MS spectra of the characteristic fragments with m/z=79.7574 (SO_3^-), m/z=96.9601 (HSO_4^-) and 122.9758 ($C_2H_3SO_4^-$) supports the proposed structures. Although ChemSpider and PubChem databases only provided linear chain candidates, ramified compounds may exist (and similar MS/MS spectra are expected). Therefore, a level of confidence 3 was assigned to these substances.

Other identified surfactants included the substances $CH_3(CH_2)_{12}(CH_2CH_2O)_4OSO_3H$ (level 3), $CH_3(CH_2)_{11}(CH_2CH_2O)_4OH$ (level 1) and $CH_3CH{=}CH(CH_2)_{15}(OCH_2CH_2)_{10}OH$ (level 3) and the additional compounds belong to the respective homologue series detected through retrospective analysis (**Figure S3-7A, S3-7B and S3-7C (SI), respectively**).

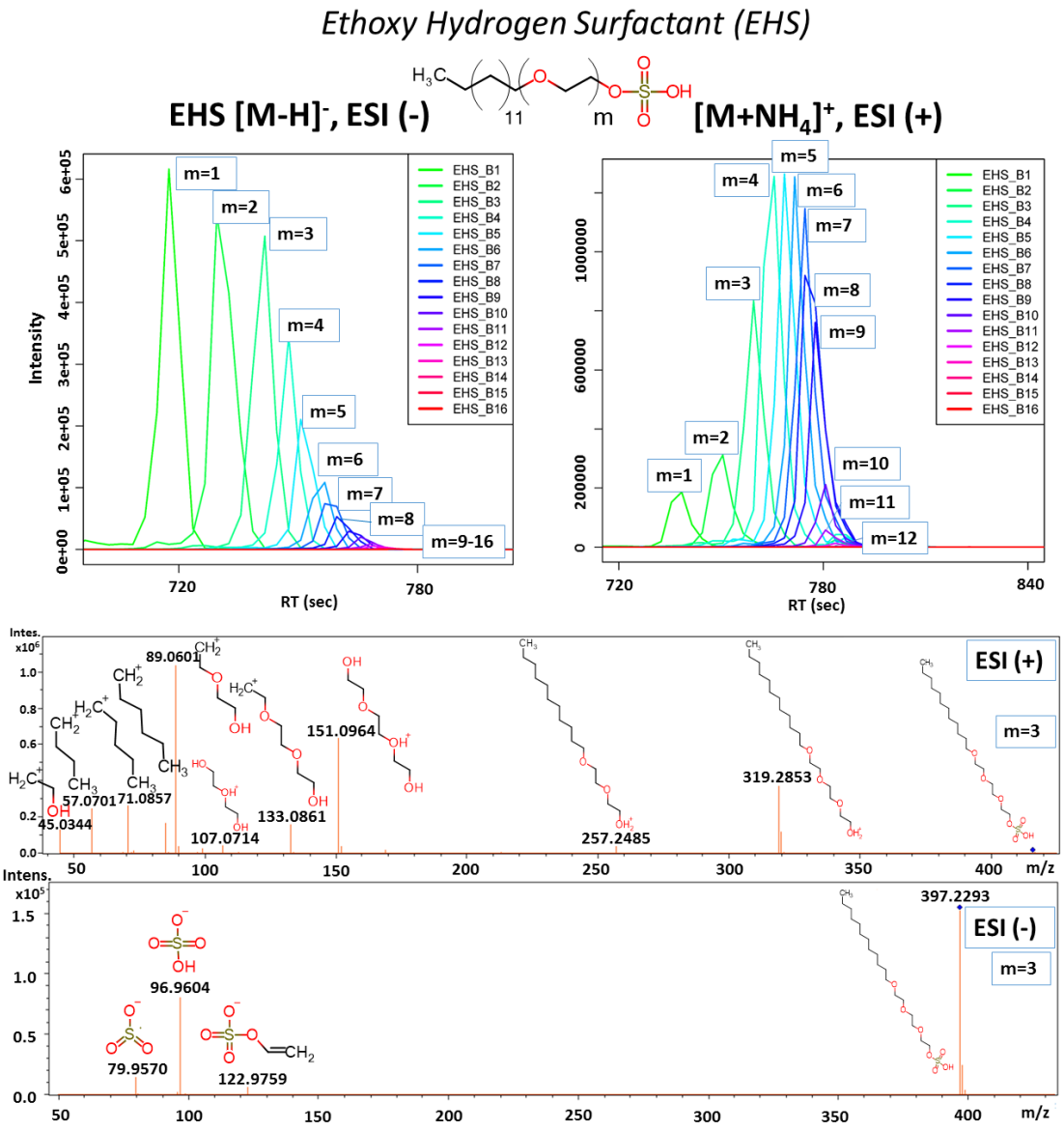


Figure 4. Tentative identification (Level 3) of a novel ethoxy hydrogen surfactant (EHS) homologue series in negative and positive ionisation mode. Spectra correspond to m=3.

In all these cases consistent t_R shifts, peak shapes and MS/MS spectra were observed. All these identifications indicate that several surfactants and their corresponding TPs remain unreported in wastewater yet.

The spectra of the successfully confirmed substances were uploaded in MassBank database (AU4064, AU4065, AU4066, AU4067) in order to assure their easy accessibility for the community of analytical environmental chemists.

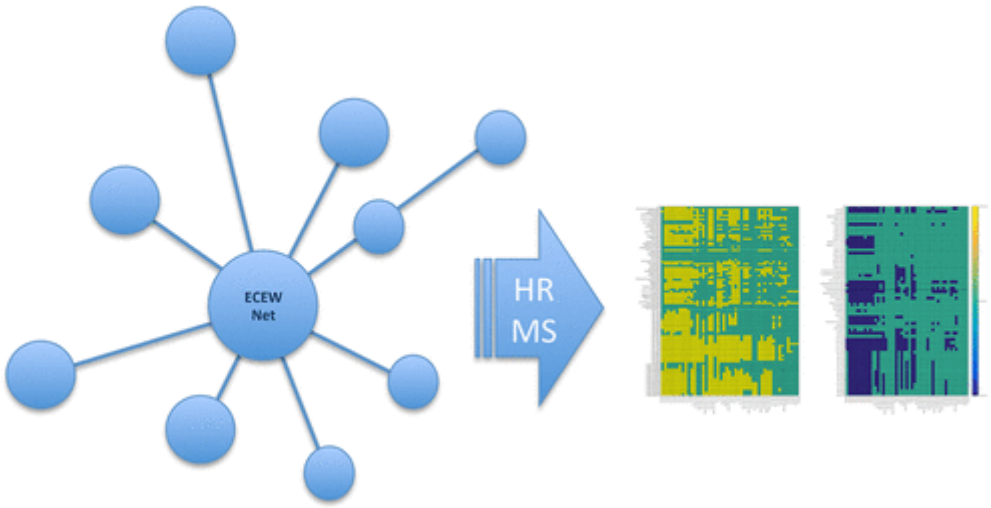
3.4 Conclusions

The developed computational workflow was successfully optimized for critical parameters as demonstrated for influent samples from the WWTP of Athens. The statistical test MEBA, which had not been used before in such identification workflows, was successfully used to prioritise compounds with large concentration variations among the samples. This success of the workflow was demonstrated by tentative identification of 14 compounds wherefrom two compounds detected for the first time in raw wastewater.

The development of new prioritisation methods capable to prioritise and identify unknown compounds in environmental samples is important as non-target screening becomes wide-spread. Smart prioritisation strategies combining the power of LC-HRMS with advanced statistics can lead to a much better understanding of the environment from a chemical point of view. However, the current lack of an interface to host the developed prioritisation approaches prevents transparent comparison of the different approaches and standardization of the methods. It also complicates the application of multiple methods to the same set of samples which may lead to the identification of an increasing number of unknown compounds. The development of unified interfaces that solve the aforementioned limitations in combination with platforms for the storage of large mass spectrometric data would provide important advances to better understand the presence and fate of micropollutants in the environment.

CHAPTER 4

EXPLORING THE POTENTIAL OF A GLOBAL EMERGING CONTAMINANT EARLY WARNING NETWORK THROUGH THE USE OF RETROSPECTIVE SUSPECT SCREENING WITH HIGH-RESOLUTION MASS SPECTROMETRY



Highlights

- Eight reference laboratories participated in the joint activity
- 48 sets of samples from 14 countries and 3 continents were investigated
- All reported results were further examined through a quality control assessment
- Samples were investigated through retrospective screening for a list of 156 analytes
- Challenges derived from the QA/QC check are discussed

This case study has been published in Environmental Science & Technology, 4th April 2018, Volume 52, Pages 5135-5144 (10.1021/acs.est.8b00365).

4.1 Introduction

One of the key challenges in the environmental and exposure sciences is to establish experimental evidence of the role of chemical exposure in human and environmental systems **[82, 83].** Our 'chemosphere' is continuously changing and most chemicals that are indexed in the Chemical Abstract Service (CAS) are not characterized with respect to their potential effects on human safety and environmental health **[84]**. Non-target analysis employing high-resolution mass spectrometers has been established over the past years as one of the key approaches for tackling this complexity. High resolution and accurate hybrid tandem mass spectrometers, such as time-of-flight and Orbitrap instruments have facilitated increased reliability in target analysis (using reference standards), enabled suspect screening (without reference standards) and screening for unknowns **[85-87]**. Substantial research effort has been placed on developing tools and workflows that expedite these three approaches, with the overall outcome that the contemporary analyst is able to obtain large amount of accurate mass data for a particular sample. For example, in 2013 the NORMAN Network organized a non-target screening collaborative trial employing target, suspect, and non-target workflows to identify substances in water samples **[81]**. This trial revealed that non-target techniques are in general substantially harmonized between practitioners and that although data processing can be time consuming and remains a major bottleneck, suspect screening approaches are very popular. However it recognized that "*better integration and connection of desired features into software packages, the exchange of target and suspect lists, and the contribution of more spectra from standard substances into (openly accessible) database"* are necessary for the technique to reach maturity **[81]**. The archiving of HRMS data also allows for data to be processed retrospectively, for example to investigate the occurrence of a newly identified compound or simply one that was not considered at the time of analysis **[28]**. This possibility has led to researchers working in this field to digitally archive data in preparation for future retrospective analysis and has even led to proposals for the establishment of data repositories, akin to environmental data banks, where digital information can be safely stored for future retrospective analysis.

Non-target HRMS full scan data allows the potential for rapid and cost-effective screening of the occurrence of newly identified contaminants in previously archived HRMS data; often referred to as retrospective analysis. Typically, it refers to the application of suspect screening workflows to archived data as reference standard measurements are not available for the analytical settings. Whilst retrospective analysis with HRMS in environmental sciences has been discussed for some time **[25, 28, 81, 88]** there are few published studies that actually apply the approach **[89, 90]**. As far as we are aware there have not been coordinated studies to investigate the spatial and temporal distribution of CECs in environmental samples through performing retrospective analysis on HRMS data acquired using different instrumental platforms and data processing software. This, has the potential to be an improved and effective strategy for establishing the extent of a newly identified contaminant's occurrence rather than the traditional approach of a new contaminant(s) being reported in the scientific literature and individual research groups subsequently validating targeted methods and reporting their own data. In order to test this hypothesis, a pilot study was performed where eight reference laboratories with available archived HRMS data were recruited with the goal of exploring the potential of a contaminant of emerging concern early warning network through the use of retrospective suspect screening employing HRMS. The pilot study was referred to as the NORMAN Early Warning System, abbreviated to NormaNEWS **[91]**.

4.2 Materials and methods

4.2.1 Participants and samples

The participants of the NormaNEWS exercise (8 reference laboratories; Eawag, KWR, NIVA, QAEHS, RWS, UJI, UoA, and Vitens) submitted samples from 14 countries and 3 continents. In total 48 sets of data from the analysis of environmental samples were evaluated. Detailed information on sample matrix, sampling date, instrument type, chromatographic separation (flow, column, gradient programs, and solvents), mass spectrometric method (acquisition mode and calibration method) are presented in the "**Sample Information**" sheet in the supporting information excel spreadsheet

(https://pubs.acs.org/doi/suppl/10.1021/acs.est.8b00365/suppl_file/es8b00365_si_002.xlsx). Further, a more detailed description of the samples and methods used are presented in the SI spreadsheet, including information on any previously published datasets.

A wide variety of environmental samples were included in this study. The majority of the samples were wastewater (effluent and influent), surface water, and groundwater samples. More than half of the samples (26 out of 48) were wastewater samples (mainly effluent wastewater samples). Wastewater sample data sets were from Switzerland (various locations) **[52]**, Norway, Sweden, Finland, Denmark, Iceland, Spain, Greece, Mexico and Australia. Fifteen of the 48 samples were samples from ecologically important large rivers such as Danube (station JDS57 Bulgarian/Romanian boarders) **[81]** and Rhine **[59]**, smaller rivers such as Swiss rivers (Furtbach and Doubs) **[92]**, Dutch rivers (Meuse and Vecht) and the Logan river in Australia. Four groundwater samples were included from Spain and the Netherlands. One primary sludge sample from the wastewater treatment plant (WWTP) in Athens (Greece) **[93]** as well as one seawater sample affected by treated wastewater [94] were also evaluated. Finally, two drinking water samples from Ridderkerk and Lekkerkerk in The Netherlands were included in the study. All the participants were asked to provide only the absolute intensity of the identified features that were blank subtracted in order to avoid the false positive identification.

Participating laboratories analysed their samples using their own routines (i.e. sample preparation and data processing) for all the analytes included in the NormaNEWS suspect list ("**NormaNEWS compounds**" sheet in the **SI**, on the NORMAN Suspect Exchange and in the CompTox Chemistry Dashboard). No specific method (i.e. chromatographic, ion source, and polarity) was recommended to the participants. This was in order to test the applicability of this approach for the data generated via different methods. For these analyses, a wide range of mass analysers as well as chromatographic conditions was employed by different participants ("**Sample Information**" sheet in the **SI**). All of the reported results were further examined, through a quality control assessment, to produce harmonized and comparable results (see section '**Quality control criteria**' in **SI** spreadsheet). Finally, each identified peak

was assigned with an appropriate confidence level **[27]**. These quality assurance steps were deemed necessary for interpretation of the results.

4.2.2 NormaNEWS suspect list

The final chemical screening suspect list consisted of 156 analytes including: 74 surfactants i.e. PEGs, C12AEO-PEGs, glycol ether sulfates (GES), linear alkylbenzyl sulfonates (LAS), sulfophenyl alkyl carboxylic acids (SPACs), and fluorosurfactants (PFAS, from several classes); 54 pharmaceuticals and their TPs (e.g. carbamazepine, carbamazepine-10-hydroxy, diltiazem, diltiazem-desacetyl, and diltiazem-N-desmethyl); 17 bisphenols; and finally 11 industrial chemicals. We considered the surfactants and the industrial chemicals as two separate families of compounds, even though a lot of surfactants may have industrial source. This distinction was made due to multiple sources for surfactants. The suspect list compounds (name, molecular formula, CAS number, SMILES, InChI and InChIKey), qualifier fragment ions and lipophilic properties (logP and log K_{OW}) are included in the SI "**NormaNEWS compounds**" sheet and are available online on the NORMAN Suspect Exchange and in the CompTox Chemistry Dashboard. The list was formed from compounds suggested by participants and typically included novel emerging substances with limited environmental occurrence as well as established widely occurring environmental contaminants (e.g. carbamazepine), which was included to assess the overall concept. A high number of the proposed substances were TPs of parent drugs that were detected through suspect and non-target screening from bio-transformation experiments. In these cases, parent drugs (e.g. citalopram and atenolol) were also included so that detection rates of the parent drugs and their TPs could be investigated. Novel surfactant compounds were also included to verify their wide-spread occurrence. In addition, the inclusion of a group of bisphenols as well as 3-nitrobenzenesulfonate, specified as an industrial chemical, were a result of non-target screening identifications. The purpose of the NormaNEWS suspect list is to provide a dynamic list of potential environmentally relevant and novel chemicals, which is enriched using expert knowledge and non-target analysis results as new data become

available. The list is available at the NORMAN Suspect List Exchange (http://www.norman-network.com/?q=node/236) and on the CompTox Chemistry Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/normanews).

4.2.3 Quality control criteria

All participants of NormaNEWS exercise were requested to submit their results together with their raw LC-HRMS chromatograms. Raw chromatograms were converted to mzML using ProteoWizard (msconvert module v.3.0.10827) **[65]**. For data acquired in data-independent acquisition mode, different collision energy channels were separated using an in-house script (available in **S4.2. (SI)**), while lock mass scans were removed. For data-dependent acquisition mode, MS/MS spectra were exported as text files (named "precursor mass_retention time") and were removed from the mzML files. Treated mzML files were converted to CDF files, which are readable from various data analysis software including Bruker DataAnalysis v.4.3. (Bruker Daltonics, Bremen, Germany), which was used here.

The performance of the following parameters was checked; mass accuracy of HRMS, stability of chromatography and presence of qualifier fragments of identified compounds in higher collision energy. A combination of an expert panel and literature information was used in order to set the threshold of each quality control criterion.

The Quality control step enabled us to minimize the effect of analyst expertise and the instrumentation on the final results given that the evaluation of the analysts and/or the instrumentation was not within the goals of this exercise. Therefore, the data points that did not meet the Quality Control criteria were excluded from the finally reported results.

4.3 Results and discussion

4.3.1 Quality control assessment

Quality control was performed to ensure that data were generated from well-calibrated instruments and that the submitted data were reliable. The first and most

important step of the procedure was to check that the mass difference between the experimental and theoretical mass did not exceed ±5 mDa, which was considered the maximum tolerable mass error in the provided complex environmental samples **[95, 96]**. This was highly relevant in assessing the confidence level assigned to each identified analyte in the list.

The mass accuracy quality control is summarized in the SI "**QC_mass accuracy_ppm/ QC_mass accuracy_Da**" excel sheet and the results presented in **Figure 5**. According to the submitted datasets, Orbitrap mass analysers showed better mass accuracy performance (absolute average mass error 0.55 mDa) comparing to other TOF instruments (absolute average mass error 0.91 mDa), based on successfully identified compounds. Mass errors are caused by the complexity of the samples, saturation of the detector (see section challenges and recommendations), and the instrument itself (i.e. the age and hardware). LC-HRMS data obtained using LTQ Orbitrap instruments showed lower mass accuracy (absolute average mass error 1.1 mDa) when compared with the LTQ Orbitrap XL (absolute average mass error 0.52 mDa), which showed lower mass accuracy in comparison with the QExactive (absolute average mass error 0.37 mDa). The method used to calibrate each instrument was also considered. LC-HRMS data obtained using a Bruker QTOF were calibrated by injecting the calibrant substance at the beginning of the chromatogram, while data from Waters QTOF (in both cases) were calibrated by lock-mass every 0.5 or 2 minutes (injecting, recording and recalibrating based on calibrant peaks appearing every 0.5/2 minutes). This resulted in the Waters QTOF performing slightly better than Bruker QTOF (absolute average mass error 0.77 mDa and 0.82 mDa respectively) according to the submitted datasets. Mass error may also occur when calibration is not considered during the mzML conversion. High mass accuracy is an extremely crucial parameter to achieve high quality results during the suspect analysis. Especially, high accuracy measurements enable a decreased number of false positive detections.

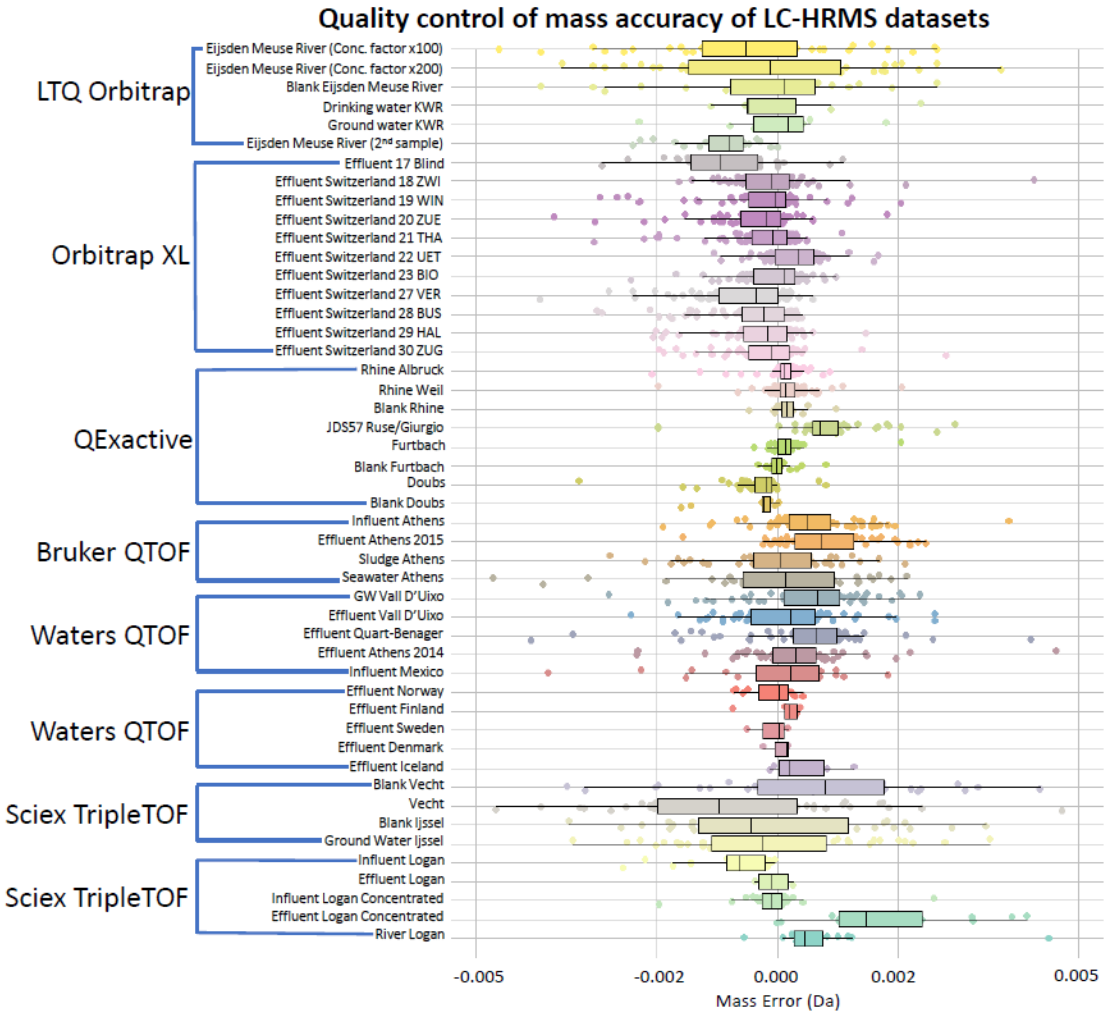


Figure 5. Quality control of mass accuracy of the submitted datasets based on the identified compounds. Type of mass analyzer, calibration type of the mass analyzer as well as other factors (age of equipment, scan sampling rate of the detector) affect the performance and the quality of the results.

The chromatographic stability of the LC separation was also assessed. All participants submitted at least 3 datasets for evaluation. Retention time data from the same instrumental set-up (and same partner) were grouped together and the normalized standard deviations (NSD) of the retention times of the detected substances were calculated (retention times of the detected substances in seconds can be found in the SI "**QC_observed_ret.time_Minutes**" sheet). A criterion of the maximum tolerable NSD of 10% was adopted for accepting the detection of a single compound across samples in data coming from the same partner. The average normalized standard

deviation of retention times in all samples was < 2% (**Figure S4-1A (SI)**). The largest variability of 8.6 % was observed for analyte valsartan, whereas the lowest variability (<0.1%) was observed for acesulfame in samples from Netherlands, GES-07 in samples from Australia, and GES-09 and GES-06 in samples from Greece. Retention time stability was considered as another extremely important parameter, which has a direct effect on the identification confidence. The low deviation observed in all the submitted datasets indicated the high quality and reliability of the LC separation of the participating laboratories.

The third QC criterion related to the presence of qualifier ions (QI) in the MS/MS spectra (**SI** "**NormaNEWS compounds**" sheet). These ions are fragments of the parent ion and are observable at higher collision energy or even at low collision energy as in-source fragments. The criterion was set on the presence of the QIs as either an in-source fragment or at higher collision energy. The identification level of compounds that did not comply with the third QC criterion were regarded as questionable and were marked accordingly **[27]**. As these QIs proved to be a very efficient way of improving the confidence of the suspect hit, Top 3 fragments have now been extracted from all mass spectra submitted to MassBank.EU and also put on the NORMAN Suspect Exchange (direct download) and the CompTox Chemistry Dashboard Downloads (direct link) for community use. The QC stage was used to exclude the features that did not meet the previously set criteria, thus harmonization. Consequently, we have reported only the features that met these mentioned criteria.

4.3.2 Overview of the retrospective screening

PolyEthylene Glycol 09 (PEG-09) was the most frequently detected compound, being present in 41 out of the 48 samples (85%) analysed. Several bisphenols, TPs of perfluorooctane sulfonate, and the pharmaceutical omeprazole were not detected in any of the samples analysed ("**Max. Absolute Intensity_counts**" sheet, **Figure 6** and **Figure S4-1C (SI)**). Series of surfactants, such as PEGs, C12AEO-PEGs, and GES, resulted in a higher detection frequency for compounds with masses varying between 400 and 600 Da compared to both smaller and larger molecules from the same families (**Figure**

S4-1B (SI)). Schymanski et al and Gago-Ferrero et al. have previously observed a similar trend for these surfactants **[23, 52]**.

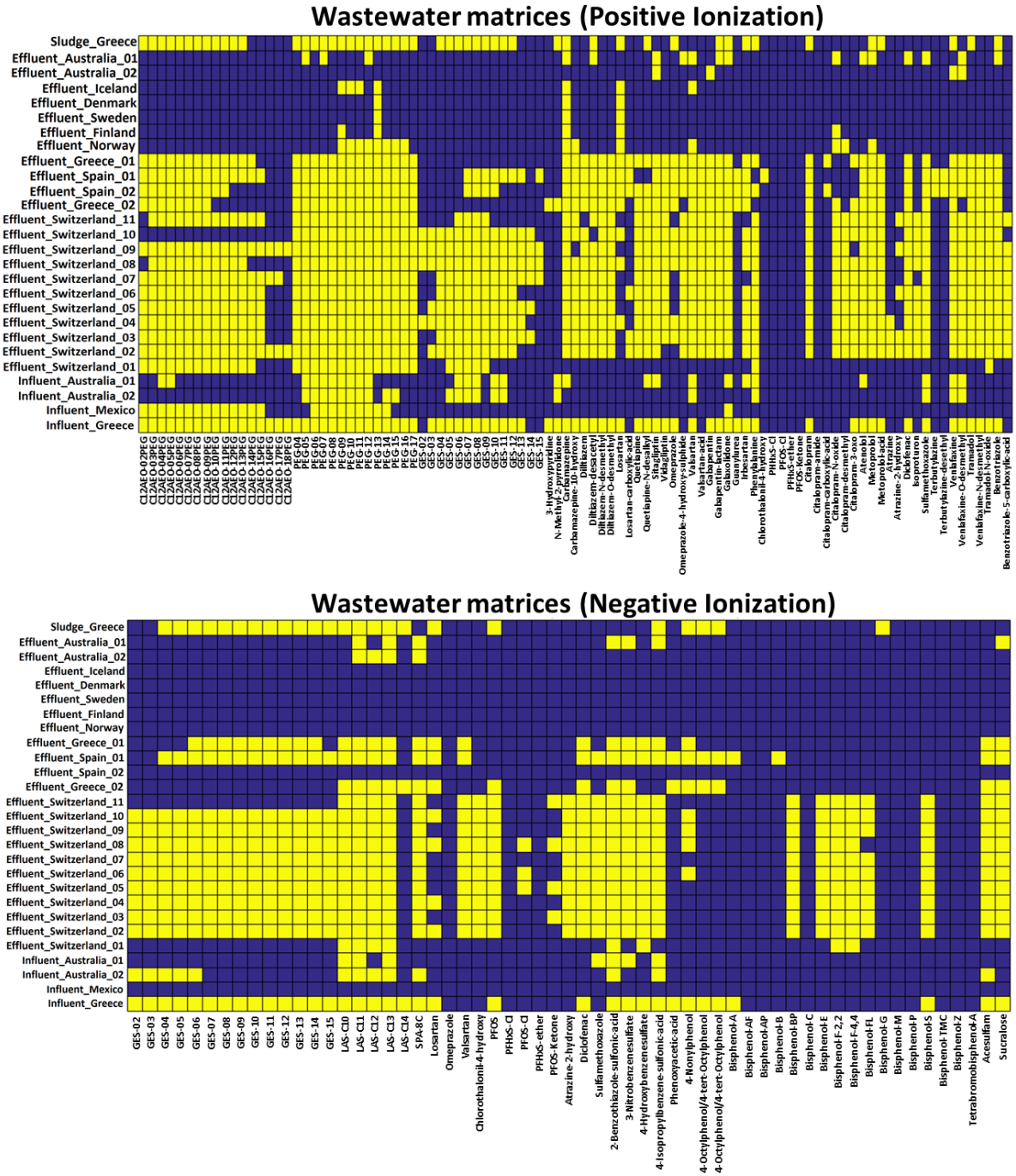


Figure 6. Heat map showing the occurrence of the selected substances in the retrospectively screened samples (primary sludge from WWTP of Athens, Greece, effluent wastewater samples from Australia, Iceland, Spain, Denmark, Sweden, Finland, Norway, Greece and Switzerland) and influent wastewater samples (Australia, Mexico, Greece) for positive and negative ionisation. Successfully identified compounds are marked in yellow.

The observed trend may be explained by the efficient ionisation of mid-size molecules compared to other compounds and potentially the fact that they are used as technical mixtures **[97]**. LAS had an average frequency of detection of around 50%. The largest measured LAS, in terms of mass (i.e. C14-LAS), were detected in only 4 samples out of 48 samples. Based on the estimated retention time for LAS-C14, we interpret that the chromatographic run times used by different partners were not sufficiently long to successfully detect this suspect analyte in the evaluated samples. Only 3 of the 5 suspect fluorinated surfactants were detected with perfluorooctane sulfonate (PFOS) having the highest detection frequency of ~ 35%. For industrial chemicals and pharmaceuticals, venlafaxine was the suspect analyte with the highest frequency of detection (68%), while several bisphenols were not detected in any of the samples. The limited data set for this pilot demonstrate the widespread presence of a number of the suspect list chemicals in different environmental compartments, particularly surface waters. At this stage further interpretation, beyond confirming the applicability of broad-scale retrospective suspect screening, should be limited to location of occurrence and frequency of detect. Future refinement of the approach will in time hopefully allow for a much more detailed assessment.

The presence of a large number of successfully detected surfactants and industrial chemicals in both wastewater influents, effluents, and surface waters suggests the wide spread these CECs in the environment across the globe, **Figure 6**. Although modern WWTPs are to some extent equipped to remove these pollutants **[98-101]**, the high production/consumption volumes of these chemicals used in households and industrial applications translates into their release into the environment. The environmental occurrence, fate and behaviour of surfactants have been widely investigated, however more reliable environmental data for these pollutants are necessary **[102-104]**. Collective exercises such as NormaNEWS are therefore an important step forward towards producing a comprehensive and reliable database on the environmental occurrence of surfactants and/or other CEC, which can be used for better understanding of their environmental fate and behaviour. Furthermore, this exercise, through the provided QC criteria, metadata template (i.e. SI spreadsheet), provides all necessary information and guidelines for laboratories across the globe for

the reliable detection, identification, and reporting of CECs in different environmental compartments.

4.3.3 Challenges and recommendations

For analysts to obtain high-confidence identifications through retrospective suspect screening they face several challenges. Here, recommendations for dealing with difficulties such as broad peaks, data acquisition, and sensitivity are provided in the following. The presence of broad peaks in the chromatograms of complex samples is often caused by the physico- chemical properties of that compound and the selected chromatographic method is unavoidable. For example, the LAS surfactants that elute at the end of the gradient of a typical reverse phase chromatographic run result in characteristic broad peaks (**Figure 7A**).

Many peak picking algorithms are unable to detect such broad peaks. Therefore, employing peak picking independent approaches **[105, 106]**, prior knowledge of those analytes, and visualization tools, even though not comprehensive, may be useful in dealing with broad peaks. The typically used data-dependent acquisition may potentially cause false identification of features due to its limitations. This acquisition mode isolates and provides MS/MS spectra of some of the most abundant ions per full scan. Even though this approach is the ideal acquisition mode during identification of peaks with the most abundant ions, this mode is not suitable for retrospective screening, due to the limited number of MS/MS spectra. In case the peak of an environmentally relevant compound is not one of those most abundant ions, the MS/MS spectra of this chemical would not be recorded (**Figure 7B**). Therefore, confident identification of that peak would not be possible. As a solution, it is highly recommended that samples are injected in data-independent acquisition mode which is the ideal acquisition mode for retrospective screening. In data-independent acquisition, HRMS is recording full scan and MS/MS spectra without prior isolation of any mass. Therefore, all fragments (and fragments of fragments in case of in-source fragments) of all co-eluting compounds are recorded, resulting in complex but full-of-information MS/MS spectra, which requires adequate data processing tool for

confident identification of features. However, to our knowledge this is the most effective acquisition method for the samples that are meant for retrospective analysis. As different compounds have different fragmentation behaviour depending on the different collision energies, the use of multiple (e.g. low, medium, high) or ramped collision energies should be considered during acquisition of data for retrospective screening to cover as many compounds as possible. As different instruments have different settings and acquisition speeds, a compromise may need to be found to provide sufficient resolution in the full scan while obtaining as much fragmentation information as possible. Pilot studies such as these and the upload of corresponding suspect lists and fragment information to public resources greatly help exchange experience to find these ideal compromises for future investigations.

Another inherent concern about LC-HRMS data is sensitivity. Among other reasons, one possible case for non-detection of pollutants is that current HRMS instruments operated in full scan are sensitive depending on the frequency with which they acquire full scans **[107]**. This means that low abundant or poorly ionized chemicals are not detected in case HRMS instrument records full scans at a high frequency rate. For example, recording full-scans at low frequency (2 Hz) will enable the detection of many compounds when comparing with a higher frequency rate (i.e. 20 Hz). Therefore, the analysts should try to find a compromise between the sampling speed and the sensitivity required for the analyses. For the samples, which are meant to be analysed via retrospective screening a lower sampling frequency is recommended given that under these conditions a higher sensitivity is achieved.

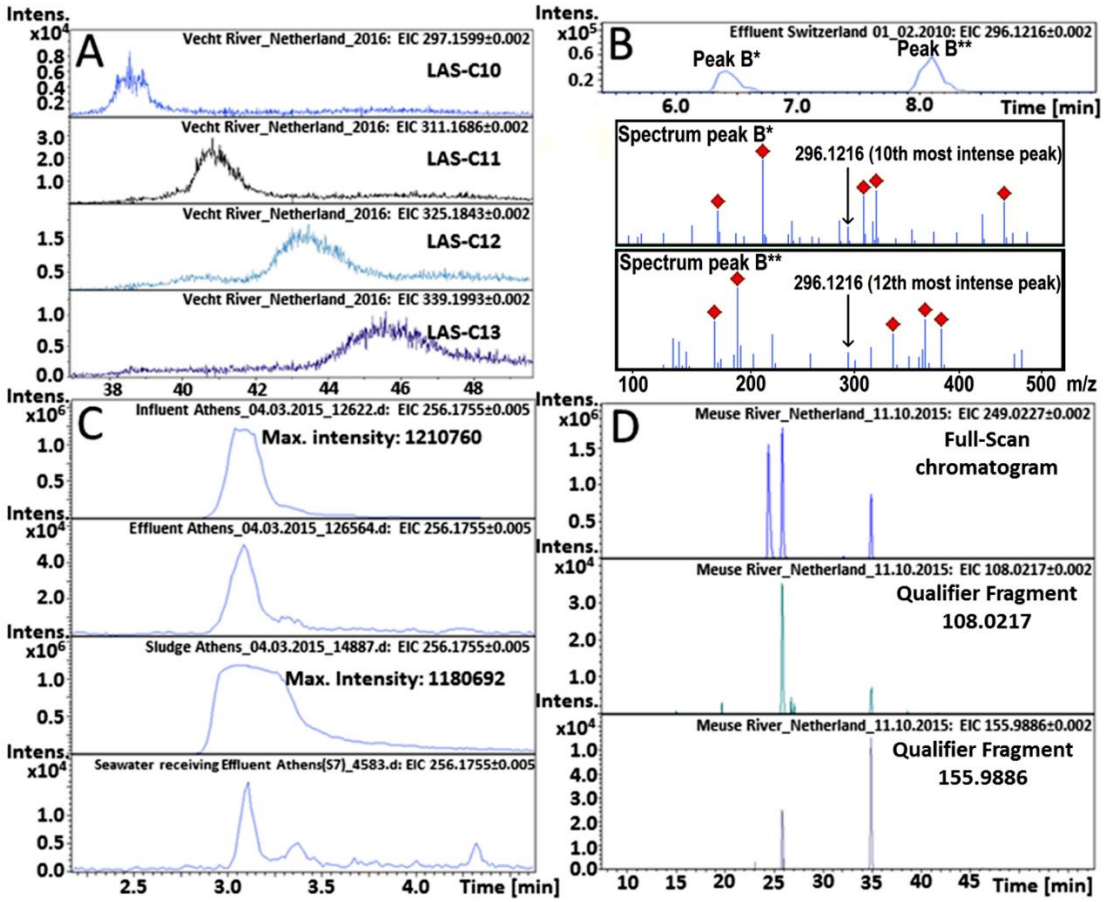


Figure 7. Challenges faced during evaluation of the results; A. Broad peaks of Linear alkylbenzene sulphonate (LAS) surfactants makes peak-picking challenging, B. Missing fragmentation information (MS/MS) of compound of interest decreases identification confidence, because data-dependent acquisition is capable to capture MS/MS only for preselected or few most abundant spectral peaks per scan (marked with red rhombus). Peaks are mass accuracy and isotopic profile consistent but not abundant enough so that MS/MS spectra have not been acquired (case of Quetiapine-N-desalkyl), C. Saturation of detector deteriorates mass accuracy, affects peak-picking and causes quantification mistakes when quantification is done by maximum intensity and not by peak area (case of PEG-05), D. Bisphenol S isomers cannot be distinguished, because in both cases qualifier fragment ions (m/z 108.0217 and 155.9886) are present in both peaks in the high collision energy channel.

Substances at high concentration levels in extracts and/or having high ionisation efficiency can often result in the detector becoming saturated (**Figure 7C**). In this case,

the peak reaches a plateau, which makes peak picking and determination of exact mass and retention time very difficult. For example, surfactants such as PEGs and C12AEO-PEGs were affected by detector saturation due to their high concentrations in the evaluated samples. The mentioned uncertainties in the exact mass and retention time are caused by the fact that saturation reduces the mass accuracy of the measurements for certain instruments, which is of extreme importance when performing identification. However, increasing the mass extraction window may solve these issues. On the other hand, such less strict mass accuracy criterion may increase the likelihood of false positive detection.

Another open issue in mass spectrometry is related to structural isomers (**Figure 7D**). Isomers are structurally similar compounds with the same molecular formula (same mass and isotopic profile) and share very similar fragmentation. This happened in case of the detection of bisphenol S in surface waters of the Netherlands. Two peaks, with different retention times, with acceptable mass accuracy, isotopic fit and same qualifier ions seem to belong to two different isomers of bisphenol S. In such cases, deeper knowledge of fragmentation behaviour and/or retention time prediction could help to identify the peak that belongs to the suspected substance. Ion ratio (ratio of the intensity of a fragment to the intensity of another fragment) can be also considered. However, this information should be carefully examined, because of ion suppression caused by high background signal produced by complex sample's matrix. Classes of substances such as the surfactants mentioned here also contain many structurally related substances that cannot be distinguished easily with mass spectrometry. These are now being grouped as "related substances" in the CompTox Chemistry Dashboard (see hyperlinks for the different surfactant classes throughout this manuscript) as a first step in working towards computational solutions to deal with the extremely complex challenge of chemical substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCBs) **[108, 109]**. Finally, all the samples need to be analysed both in positive and negative mode in order to cover a wider chemical space compared to only single polarity.

4.3.4 The early warning system and its potential

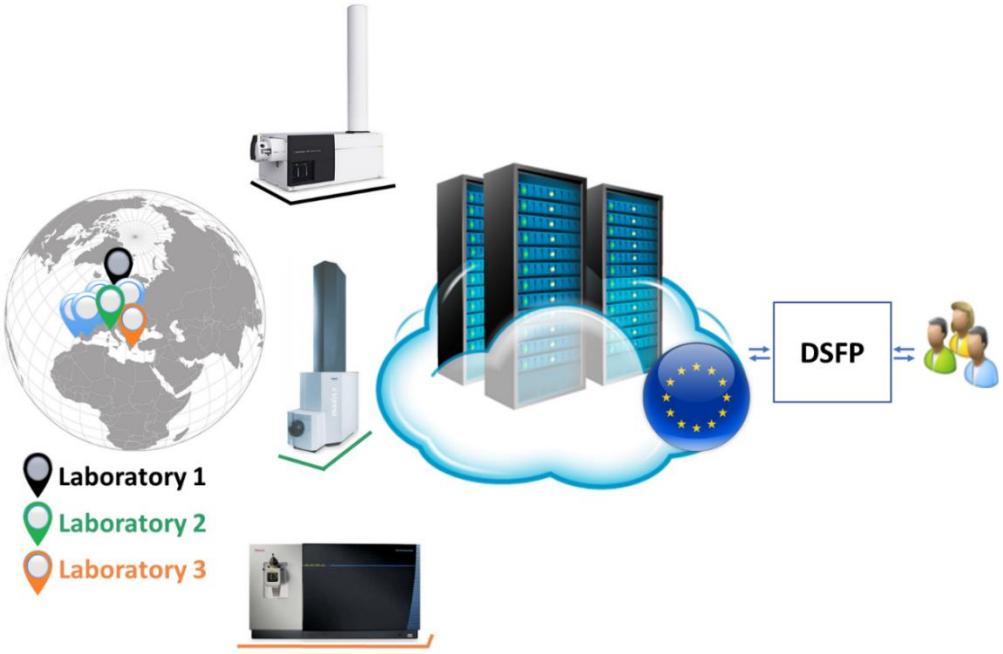
This exercise confirmed the high occurrence frequency of several surfactants (e.g. PEGs and C12AEO-PEGs), TPs of selected drugs (e.g. gabapentin-lactam, metoprolol-acid, carbamazepine-10-hydroxy, omeprazole-4-hydroxy-sulphide, 2-benzothiazole-sulfonic-acid), and industrial chemicals such as 3-nitrobenzenesulfonate and bisphenol-S. These chemicals are not typically included in target/suspect lists used for surface water monitoring programs. Subsequently, there are limited environmental occurrence data available for these pollutants **[110-112]**. This clearly demonstrates that an early warning network such as NormaNEWS enables the efficient and reliable detection and identification of novel CECs in different environmental compartments at both a temporal and spatial scale. Consequently, a reasonably large and diverse dataset on the environmental occurrence of novel CECs in different matrices has been generated for this pilot project. Clearly, this study was a proof of concept to test the applicability of such an approach to a diverse global dataset. Further development and significantly larger global coverage is necessary in order to generate a dataset suitable for both environmental interpretation and policy making practices. Such a dataset provides an initial screen that can be used to inform contaminant prioritisation exercises leading to further monitoring, fate and effect studies and subsequent risk assessment. Furthermore, given that the data are harmonized across a large number of laboratories and the confidence level of each identification is provided, the inherent reliability of each identification becomes more intuitive to non-experts. The purpose of this network activity would not be to replace ongoing targeted monitoring and screening programs, but to provide a robust and comprehensive complementary collaborative approach for informing the refinement of priority substance lists. This also shows the great potential for screening much larger lists in the future, although the manual verification of the results is still a demanding task. More computationally efficient methods will be needed before this can be expanded to potentially lists of tens of thousands of substances.

The NormaNEWS pilot was performed using a very simple approach where all participants manually submitted data on their CECs of interest in order to create a suspect screening list for the collaborative exercise. This enabled researchers to easily

obtain additional data on the CECs that they are particularly interested in. Future lists could be generated by a number of different approaches including from open resources, such as massbank.eu. As highlighted recently by Schymanski and Williams **[108]**, open resources will be instrumental in defining the evolution of suspect screening. The community-wide sharing of CECs through the exchange of suspect lists (e.g. the NORMAN Suspect Exchange and the Chemistry Dashboard lists) as well as tentatively and unequivocally identified spectra and sharing the related fragments is therefore key to the success of a global early warning network. Also key will be the willingness of the scientific community to share their HRMS data in an open MS format (e.g. mzML **[113]**, mzXML **[114]**, and netCDF **[115]**). The Global Natural Products Social Molecular Networking (GNPS; http://gnps.ucsd.edu/) provides a vision as to how global collaboration and social cooperation can be used to address major scientific challenges in the sharing and community curation of MS data **[116]**. Taking inspiration from GNPS, we propose that HRMS data are made available (through a virtual repository and with necessary metadata) in order to facilitate living data along with periodic automated re-analysis of data (i.e. with updates to the suspect list or the addition of new data sets). Ideally, this repository will be easily accessible through a web-application and free of the aforementioned challenges. The environmental and exposure sciences currently lag behind other fields, such as proteomics **[117]**, metabolomics **[118]** and natural product research **[119]** in globally collaborating and sharing data through open/social platforms in order to revolutionize the way data are processed to achieve significant outcomes. We acknowledge that not all the data tools are currently in place to make our proposal a reality, however progress is being made in this area **[105, 106, 120]**. For example, within the NORMAN Network (http://www.norman-network.net/) there is an initiative to develop a digital sample freezing platform. A global emerging contaminant early warning network based on adopting the successful practices of other similar networks will play a pivotal role in identifying chemicals using HRMS that has the potential to possess significant outcomes in protecting human and environmental health.

CHAPTER 5

NORMAN DIGITAL SAMPLE FREEZING PLATFORM: A EUROPEAN VIRTUAL PLATFORM TO EXCHANGE LIQUID CHROMATOGRAPHY HIGH RESOLUTION-MASS SPECTROMETRY DATA AND SCREEN SUSPECTS IN "DIGITALLY FROZEN" ENVIRONMENTAL SAMPLES



Highlights

- An overview of the current HRMS screening steps is presented in an automatized tool
- Digital archiving and retrospective suspect screening were integrated in DSFP
- DSFP enables enhanced data mining and visualization capabilities
- 54 seawater, 19 sediment and 12 biota samples were screened for 1447 substances
- 80 chemicals from REACH registry and 12 antibiotics were detected

This case study has been accepted for publication in Trends in Analytical Chemistry

5.1 Introduction

Tens of thousands of chemical substances are produced in Europe and worldwide in large amounts with potential to enter the environment **[16, 24]**. Many of these substances and their TPs are potentially toxic to flora, fauna and humans, but the scientific tools to establish a critical mass of evidence for this ever increasing chemical impact to support the respective legal regulatory tools are not yet sufficiently in place **[24]**. The NORMAN network, as an interface between science and policy-making, has been working to establish a retrospective system able to detect any contaminants of environmental concern (CECs) that may be harmful to environmental or human health for over a decade **[121]**. In 2017, over 70 NORMAN members in Europe and North America decided to expand from considering only hundreds to tens of thousands substances in their activities. A domain of NORMAN suspected pollutants was established **[122]**, evidence of their occurrence in the environment was collected **[123]**, and a toxicity threshold was assigned to each of the substances **[124]**. This information was intended for use in identifying/prioritising compounds (exceeding the toxic threshold values at many sites) that should be considered for regulation in Europe **[125]**.

HRMS instruments such as quadrupole-time-of-flight (Q-TOF) and hybrid quadrupole or ion trap Orbitraps acquire accurate mass and high-resolution MS and MS/MS (or in some cases MS^n) full-scan spectra and can be used to perform comparative sample evaluation **[81]**. In the margins of the analytical conditions (e.g. sampling, enrichment method and solvents used) and instrumental limitations (e.g. ionizability, selectivity, sensitivity and resolution), the full mass spectral information of detectable, suspected and unknown compounds is stored in raw data files. However often the measurements are only evaluated partially (e.g. for the targeted analysis of a given list of CECs) and most of the detected peaks remain unannotated and thus unknown **[126].** One main reason why stored HRMS data are still underexploited is the lack of software tools for data archiving, quality control and exchange **[24, 127]**. This reduces the potential for use of such data in regulation and thus limits the general ability to perform in-depth investigations into environmental contamination. The second main reason is due to a general lack of large LC-HRMS/MS mass spectral libraries and the

inability of current libraries to cover all suspected CECs and their TPs **[49]**. The NORMAN MassBank (https://massbank.eu/MassBank), established by the NORMAN network in 2011 **[128]**, is currently populated with 48,822 experimental mass spectra on more than 10,000 substances **[129]**, while other public HRMS libraries include up to a few tens of thousands substances, with some overlap among them **[49, 130]**. The potential of a public mass spectral platform for raw data to search for the non-regulated emerging substances through the use of retrospective suspect screening with high-resolution mass spectrometry was first tested in NormaNEWS pilot study **[38, 131],** in which laboratories around Europe were asked to check a pre-defined list of newly-identified CECs (https://comptox.epa.gov/dashboard/chemical_lists/normanews) in mass chromatograms of environmental samples stored locally in participating laboratories.

In this critical discussion article, the NORMAN Digital Sample Freezing Platform (DSFP) is presented for archiving, processing, analysing, data mining and retrieving information from the large amount of environmental mass spectral information derived by the community of environmental scientists and deposited at NORMAN. DSFP incorporates all the latest developments in HRMS screening **[2]** and offers an integrated tool for wide-scope screening of CECs in the environment. The primary intended uses are retrospective analysis of newly emerging substances and comparison of mass spectral data across compartments (e.g. water, biota, sediment, indoor environment, air), however additional uses and exploration of other potential applications are strongly encouraged. For example, the outcomes can be used to indicate the occurrence and spatial distribution of a certain substance within a geographical region (country, river basin, pan-European scale, etc.) or to prioritise unidentified features or compounds for future identification efforts. DSFP promotes automation of retrospective screening, enhances the transparency of LC-HRMS data and serves as a tool for drafting future policy recommendations related to chemicals management in the environment. It was tested using samples from the Joint Black Sea Survey (JBSS) covering seawater, sediment and biota matrices **[132]**. All JBSS samples were investigated retrospectively for a list of 670 antibiotics (list S6, ITNANTIBIOTIC, at https://www.norman-network.com/?q=node/236). Antibiotics were selected due

to their specific adverse effects on flora and fauna, which are exhibited even at low concentrations **[94, 133].** Their presence in the environment may also induce AR **[134]**, which has been identified as one of major global threats to the society **[135]**. Additionally, samples from the JBSS were screened for 777 compounds extracted from the REACH registry in cooperation with the ECHA in 2017.

5.2 Databases

5.2.1 Candidate substances

The NORMAN Network (www.norman-network.com) is a unique network of reference laboratories, research centres and related organisations for monitoring of CECs in the environment. A working group (WG) on non-target screening (NTS) was established in 2015 with the aim to promote and integrate all the research efforts in NTS of environmental samples within Europe and beyond. One of the main efforts of the NTS WG was the development of DSFP. DSFP was designed for archiving LC-HRMS chromatograms and for the retrospective screening of polar and semi-polar CECs in various environmental matrices. Such methods should include a generic sample preparation methods, such as those presented for wastewater **[64]**, sludge and sediments **[136]** and biota **[29]**. A recent collaborative trial revealed that suspect screening was a very common and efficient way to expedite non-target screening **[81]**. As a result, the NORMAN Suspect Exchange initiative was founded (http://www.norman-network.com/?q=node/236) and members were encouraged to submit their suspect lists. To date, 40,053 highly varying substances have been collected and curated. All preparation and curation was done within the network using open access cheminformatics toolkits and formed one merged list (SusDat) **[122].** Antibiotics and their major TPs is part of SusDat and the compounds were mined from the literature, compound databases and chemical catalogues. A list of the reviewed sources can be found in **S5.2 (SI)**. The list of 670 antibiotics (termed "ITNANTIBIOTIC") is accessible online at NORMAN Suspect List Exchange **[137]**, which contains environmentally relevant substances collected and contributed by NORMAN members. The literature review showed that only a small fraction of the collected

antibiotics (66 out of 670) was mentioned in the literature (**Table S5-2A**), whereas for the rest of antibiotics occurrence data in environmental samples is rather scarce (**Figure S5-2A and S5-2B (SI)**). The spectral libraries m/z Cloud **[138]** and NORMAN MassBank **[128, 129]** were queried for these substances. Experimental mass spectra, including information on preferable ionisation of the compound, screened adduct and qualifier fragment ions were found for 159 antibiotics. The rest of substances were screened assuming the presence of $[M+H]^+$ and $[M-H]^-$ ions using *in silico* predicted fragments calculated using CFM-ID **[30]**.

5.2.2 Data collection templates

Each sample was stored digitally in DSFP as a separate Excel spreadsheet-based "Data Collection Template" (DCT), which is linked to one or more raw mass spectral files (mzML file format) according to the number of different HRMS methods applied for analysis of the sample. Each DCT consists of six sections; organization details, retention time index (RTI) information, sample description, analysis, instrumental metadata and fragment peak list (https://norman-data.eu/DCT_NTS.xlsx). This DCT was used to report non-target screening (NTS) results from reference laboratories participating in collaborative trials organized by NORMAN in 2015 and 2016 **[17, 81]**. Analysis of sample extracts by LC-HRMS result in binary data, which have different structure and format, so that files are accessible for processing by the respective vendor's software. To allow for interoperability among results obtained by instruments from different vendors, files are converted to a commonly agreed format (mzML) **[113]** by converters, among which the most widespread is ProteoWizard **[65]**. For exporting JBSS chromatograms in mzML format in this study, CompassXport 3.0.9.2. (Bruker Daltonics, Bremen, Germany) was used.

5.2.3 Repository

To acquire suitable LC-HRMS data and use DSFP to its full potential, the procedure described in section **S5.3 (SI)** should be followed. The procedure involves actions to

be taken before, during and after the instrumental analysis. For future import of new samples, contributors should follow the same scheme and, while the choice of mobile phases, gradients and reversed-phase columns can be according to their protocol, they should (i) assure that their equipment is clean and well-calibrated, (ii) inject RTI mixture during the sequence (**Figure S5-3E (SI)**), and (iii) analyse sample extracts in data-independent and data-dependent modes, as explained in **Figure S5-3F (SI)** for optimal results.

The R-based workflow for importing new samples into DSFP is represented in **Figure 8**. LC-HRMS chromatograms converted to mzML can be processed with community driven workflows **[113].** There are many workflows available, but the final output is always a component list **[67, 117, 139]**. DSFP uses the centWave algorithm (through xcms R-package v. 3.4.1) for peak picking with previously optimized *ppm* and *peakwidth* parameters **[75, 76]**. The peak picking algorithm searches for consecutive masses within a mass error threshold forming peak shape in chromatographic dimension **[67]**. The next step is componentization, which is a procedure for grouping peaks coming from the same compound (adducts, isotopic peaks, in-source fragments) **[68, 140, 141]**. For this purpose, functions from nontarget R-package (v. 1.9) were used **[140]**.

The next step involves calculation of an experimental Retention Time Index (RTI) for every feature detected based on the retention time observed and the calibration curve equation (RT=f(RTI)), derived from the retention time of the standard calibration mixture **[142]**. This is followed by the extraction of HRMS/MS fragments that were isolated and fragmented, in case the sample was analysed using a data-dependent method. Finally, the output is shaped to the DCT format after adding organization details, information on sampling site/date/matrix and instrumental metadata. The chromatograms are contributed *via* a web interface (**Figure 5-3D (SI)**), which automatically generates a DCT for each sample and facilitates the upload of the respective sample-specific mzML files to the server in a harmonised format ready for processing.

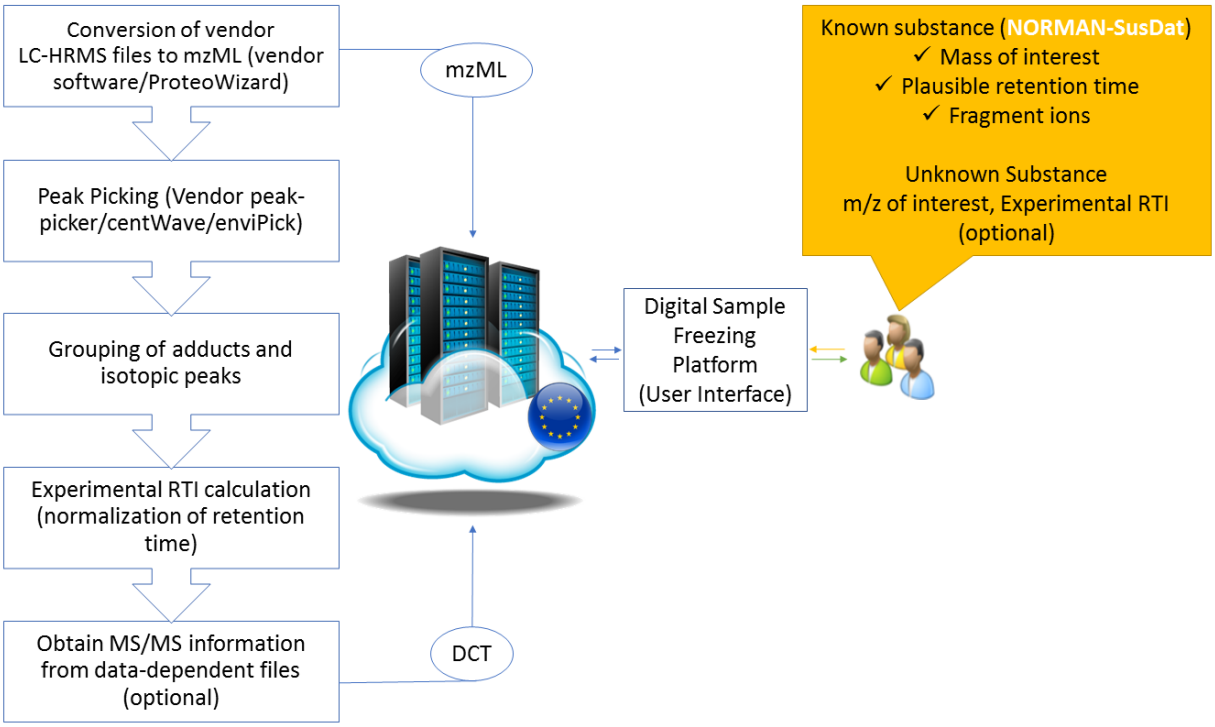


Figure 8. Adopted workflow for obtaining harmonised mzML raw data formats (provided by users) and automatically generated Data Collection Templates (DCTs) accessible to users through the Digital Sample Freezing Platform (DSFP) interface.

5.2.4 Overview of the screening process

Once DCTs and mzML files are imported into DSFP, the user may search these and/or other samples for the presence of a single substance (**Figure S5-4A, (SI)**) or for many compounds included in SusDat (40,053 suspected CECs as of November 2018) **[122]**. Individual searches for unknown compounds based on their exact mass is also available. SusDat contains important information for the screening of CECs in DSFP (exact mass of a molecular adduct ions $[M+H]^+$ and $[M-H]^-$, predicted RTI) supplemented by masses of experimentally observed or predicted fragment ions. If no experimental fragments were available in MassBank for a given compound, then *in silico* predicted fragments were used instead, calculated with CMF-ID **[30]**. The fragmentation pattern of all NORMAN MassBank compounds is integrated to SusDat (list S1 "MASSBANK" at https://www.norman-nework.com/?q=node/236). Therefore, the platform automatically suggests to search for the exact mass of a compound using

a preferred specific ionisation mode (positive or negative or both) and an expected adduct (like $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ in positive ionisation mode) according to the available experimental data. For tentative identification (indicated by summary of identification evidences), the platform considers mass accuracy, plausible (window of) retention time in the chromatogram, isotopic fit and a presence of matching fragments. Unknown substances not included in SusDat can also be searched for occurrence over multiple samples by exact mass. In all cases, the user must specify the mass accuracy error, the tolerance in plausibility of retention time (by default values are proposed) and fragments.

Compound querying is a rather straightforward procedure, in which user selects (i) a compound of interest from a drop-down list of SusDat substances and (ii) samples in which the compound should be searched. Compound selection is possible using any of the following identifiers: common name, CAS number or InChIKey **(Figure S5-4B (SI)**). Mass chromatograms can be filtered down to those desired to be submitted for investigation based on the country, matrix type and project (**Figure S4-4C (SI)**). Then a search can start, in which the exact mass of the ionized form of the substance is searched for a match in the DCTs of all selected mass chromatograms. Features that pass mass accuracy and fit into the expected RTI window are presented in an interactive and downloadable table (**Figure S5-4D (SI)**). A column with detected fragment ions is presented in a format “exact mass/retention time/absolute maximum intensity”, which allows for a quick check whether the proposed identification is plausible (**Figure S5-4D (SI)**). DSFP offers the possibility to change parameters such as mass accuracy tolerance or RTI tolerance (**Figure S5-4A (SI)**). As an example, RTI tolerance could be set to 100% in situations where one does not wish to consider the RTI values and instead rely only on exact mass ions-based identification.

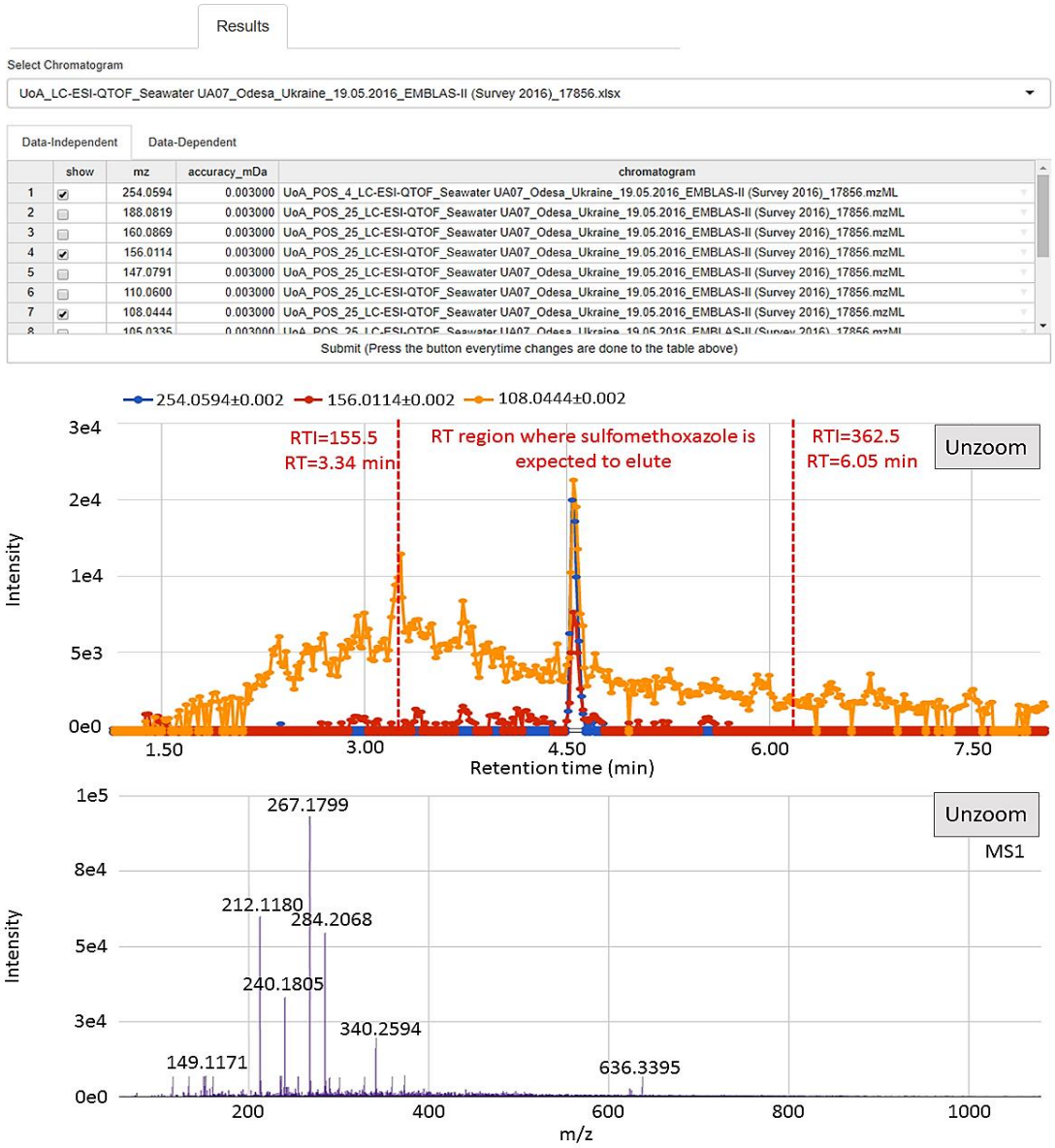


Figure 9. Extracted ion chromatograms (*m/z* $[M+H]^+$ = 254.0504, fragments from MassBank *m/z* = 156.0144, *m/z* = 108.0444) indicating the presence of sulfamethoxazole in the Black Sea water sample from the area close to the delta of the Danube river. The molecular ion and two structure characteristic fragments support the identification of the structure, the RTI window from 155.5 to 362.5 serves as complementary supporting evidence (Level 3). The compound was eventually confirmed with a reference standard (Level 1).

From the results table, it is possible to go back to the raw mass chromatograms stored in the repository and perform extracted ion chromatogram (EIC) search and/or have

a look at the full scan spectra of the selected chromatographic peak (**Figure 9**). DSFP will automatically retrieve fragments for compounds with experimental spectral information available in SusDat (2219 compounds as of December 2018) and add them in an interactive table to help users verify the identity of the compound they searched. For example, in **Figure 9** the identity of the antibiotic sulfamethoxazole was supported by the presence of molecular ion adduct and two matching fragments. The EIC table is interactive*, i.e.* rows with different exact masses of interest can be added manually and thus visualized on the screen, whereas mass accuracy and mass chromatograms selected for screening can be changed according to the choice of the user.

5.2.5 Investigation of spatial distribution of detected compounds

The results of the search can also be visualized in an interactive map (**Figure 10**). Observed intensities of the compounds are normalized based on the maximum observed intensity over all samples. Moreover, the enrichment factor is also considered in case the sample was enriched, e.g. by solid phase extraction (this information is mandatory during the contribution procedure). Finally, the intensities of a compound are shown on the map as scaled circles; the bigger the size and the more intense the colour - the higher the signal and presumably concentration of the substance. The scaled mapping of the normalised intensities allows for a user-friendly visual assessment of the spatial distribution of substances of interest and possible sources of pollution.

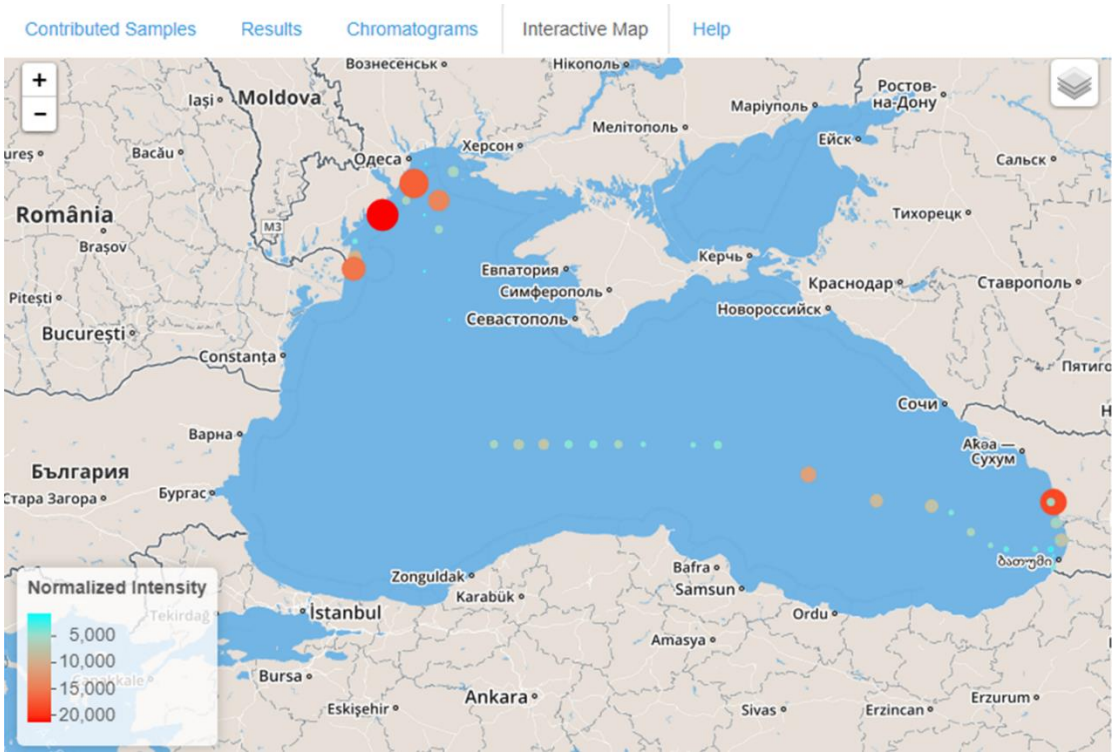


Figure 10. Spatial distribution of DEET in seawater samples from the Black Sea. The profile reveals that there is land input of DEET to the Black Sea. DEET was detected in 39 (out of 54) water samples analysed by LC-HRMS and using the NTS workflow (for details, see Methodology). The circles indicate sampling stations; small, light blue circles relate to low normalized intensities; large red circles indicate high normalized intensities of DEET.

Figure 10 shows the example of DEET, the spatial distribution of which suggests inputs from various diffuse sources. This visualization method is suitable for results obtained for samples using the same sample preparation and analytical conditions and coming from the same instrument but may not be accurate for comparing samples coming from different instruments. Nevertheless, even in this case, it will give a quick rough overview on the presence of investigated chemicals in different locations. As a further example, in **Figure S5-4E (SI)**, the highest normalised intensities of sulfamethoxazole were detected in three Black Sea sampling stations close to the Danube delta during JBSS 2016 **[132]**, suggesting the Danube as a pollution source. This is in line with observation obtained in the Joint Danube Survey 3, conducted earlier in 2013, where sulfamethoxazole was reported at relatively high concentration levels 15 and 16 ng L^{-}

[1] in sampling stations Sf. Gheorghe arm and Vilkovo, respectively, in the areas close to the above Black Sea sampling stations **[143]**.

5.2.6 Batch mode module and interactive heatmap visualization

The batch mode module provides the possibility to search for up to thousands of compounds included in SusDat in a single batch in all or a selected number of samples stored in DSFP. Again, the user specifies compounds and samples to be investigated and obtains detailed and summarized results by a single click of 'Submit' button (**Figure S5-4G (SI)**). DSFP returns a summary spreadsheet file containing absolute maximum intensity of the observed signals, experimental retention time, mass error/accuracy (in mDa and ppm) and information on detected fragments including exact mass/retention time/absolute maximum intensity (**Figure S5-5**). For each analyte there is a common name, molecular formula, CAS No., SMILES, InChI, InChIKey (all retrieved from SusDat), a column with the identification evidence (*i.e.* mass accuracy, isotopic fit, plausibility of retention time and number of fragments) and a column whether fragments are predicted *in silico* or obtained experimentally (extracted from available HRMS libraries). A detailed report is provided in a format of multi-sheet spreadsheet file, in which each sheet represents one compound with the same content as obtained from the single compound search (**Figure S5-5f**). The total processing time depends on the number of selected compounds, samples and computational power. Typical processing time for screening of 1000 suspects in 86 JBSS samples was 13±1 min for an Intel® Core™ i7-4702MQ CPU processor at 2.20 GHz.

The batch mode tool is equipped with its separate interactive graphical presentation tool - a heatmap, such as the one presented in **Figure 11** and **Figure S5-5B (SI)**. Here, the selected compounds are presented in rows and samples in columns. White colour in the heatmap means that compound was not detected in the sample, whereas blue means positive detection. In a simplified scoring system, compounds that satisfy the mass accuracy criterium and have a plausible retention time (observed within the RTI window) receive one point and one additional point for each matching fragment ion. As shown in **Figure S5-5B (SI)**, the user can customize which compounds appear in the

heatmap (compounds with predicted or experimental fragments, compounds that exceed a specified score), as well as the appearance of heatmap.

In this study, an identification was considered as having substantial supporting evidence if three or more fragments were detected for compounds with library spectra available and if five or more fragments were detected for compounds with *in silico* predicted mass spectra, in addition a match of exact mass of the molecular ion and a plausible RTI. The number of fragments is critical to distinguish between false positive and false negative identifications. It should be made clear here that DSFP is providing supporting evidence for tentative identification and does not aim at assigning identification levels. While all proposed identifications remain technically as a Level 3 **[27]**, those with substantial supporting evidence are clearly higher confidence as those with only an exact mass match. All plausible identifications should be verified with an exact library match (to obtain a Level 2a status) or confirmation of retention time and fragment information with a reference standard for Level 1. While DSFP is not able to perform this directly at this stage, instead DSFP offers the users an opportunity to further explore the data by going back to raw mass chromatograms by performing extracted ion chromatograms and by visualizing the MS^n spectra (e.g. as shown in **Figure 9**).

5.2.7 Screening of antibiotics in the Black Sea samples

Before the application of DSFP screening in real case studies, the results were benchmarked against the results of a target screening method. The JBSS samples were subjected to wide-scope target screening of 2248 compounds (list "UATHTARGETS" **[137, 144]**). The validation results are summarized in section **S5-6 (SI)**. DSFP was used to screen antibiotics in JBSS samples using the "ITNANTIBIOTIC" list. Twelve out of 670 suspect antibiotics were detected. Following further evaluation of DSFP results, eight compounds were confirmed by analysis of reference standard at level 1, three were identified at level 2a (probable structure by library spectrum match) and one at level 3 (tentative candidate, see **Table S5-7**) **[27]**. Antibiotics with the highest frequency of detection were sulfamethoxazole (44 out of 86 samples, mainly in seawater),

aminosalicylic acid (34 out of 86 JBSS samples), chlorhexidine (15 out of 86 samples, mainly in sediment) and 8-hydroxychinolin (11 out of 86 samples, mainly in biota), while the other nine compounds were only detected in few samples. These last nine antibiotics detected sporadically were mainly found in seawater samples close to the Danube delta. Macrolide monensin, griseofulvin, lincomycin and the sulfonamides sulfadiazine and sulfapyridine, all confirmed at level 1, were detected in the sample from the Ukrainian shelf (Sampling station UA07; **[132]**) and some of them even at stations more distant from the Danube delta (Sampling stations UA06, UA05). Sulfadiazine was also detected on the Georgian coastline together with fluconazole, the presence of which was confirmed from mass chromatograms obtained in both positive and negative ionisation (**Figure 11**).

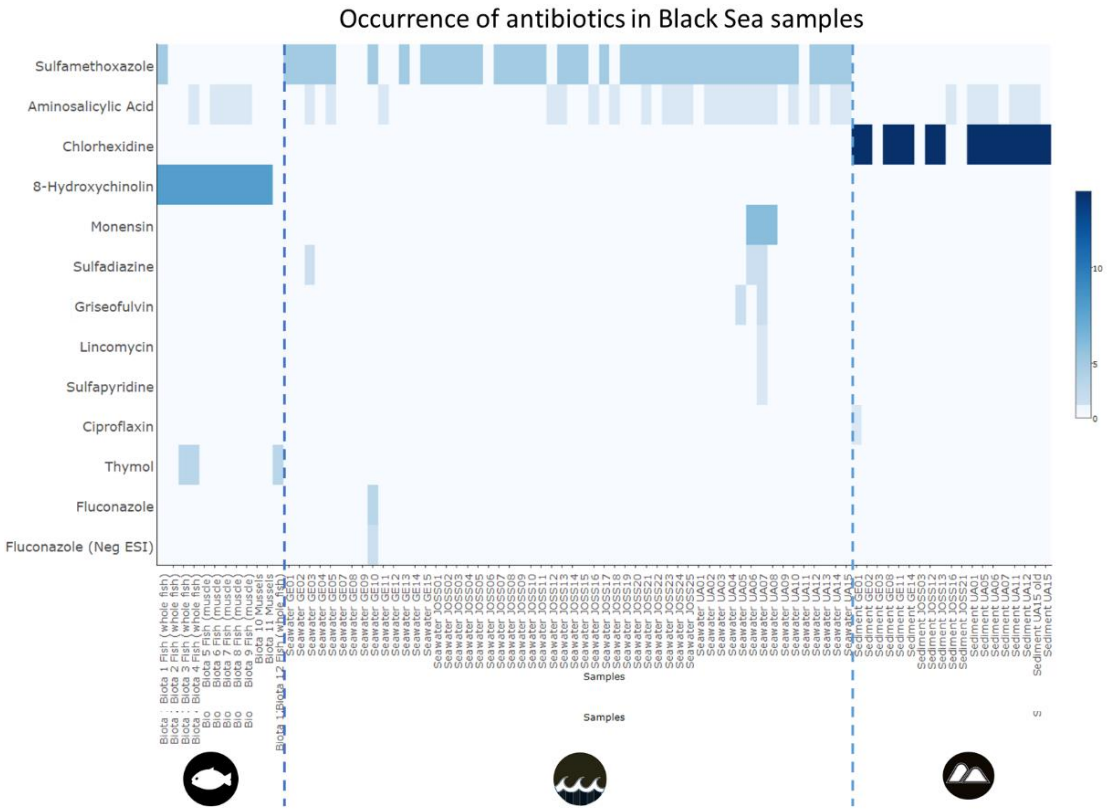


Figure 11. Occurrence of antibiotics in the Black Sea samples collected within the JBSS 2016 **[132]**. Blue coloured rectangles indicate that a compound was detected in the sample; the darker the colour, the more structure-related fragments were detected in the corresponding mass spectra. The first 12 samples (x-axis) are fish and molluscs, the following 54 samples are seawater followed by 19 sediments samples. The bar on the right-hand side indicates colour code related to the score of identification.

Compounds are ordered by increasing frequency of appearance. For more details, see text.

Chlorhexidine was detected in 15 out of 19 sediment samples, including samples taken from the seabed of the Black Sea at a depth of more than 2000 m, confirming its widespread occurrence and persistence. Another compound with widespread occurrence was aminosalicylic acid (level 1), which was detected in seven out of 19 samples. Thymol, 8-hydroxychinolin, aminosalicylic acid and sulfamethoxazole were detected in biota samples. Occurrence of thymol is not considered of particular concern at this stage since it is a naturally occurring compound with antimicrobial and antibiotic properties. However, 8-hydroxychinolin has a wide range of applications and a tendency for bioaccumulation and persistence in the environment **[145]**. The presence of this compound in all tested biota sample should be further explored. Aminosalicylic acid, an antibiotic primarily used to treat tuberculosis, was detected in five out of twelve biota samples. The presence of this substance in all three investigated matrices (sea water, sediment and biota) is potentially alarming. Sulfamethoxazole was detected also in one biota sample and its presence in the marine ecosystem deserves attention.

Only 12 out of 670 antibiotics were detected in the investigated samples. However, occurrence of these antibiotics in such a diluted matrix as sea water and their accumulation in sediments and biota far from the sources of pollution should draw the attention of both ecologists and regulators and demonstrates the potential for DSFP to provide tentative information about chemicals of concern in certain areas and reveals the power of mass spectrometry to help find unexpected contamination in the environment.

5.2.8 Screen of REACH compounds in the Black Sea samples

Out of ca. 68,000 substances that can be found on the official website of ECHA containing registered REACH compounds (https://echa.europa.eu/information-on-chemicals/registered-substances), only 777 compounds had experimentally observed

mass spectra stored in the NORMAN MassBank. The lack of coverage of REACH chemicals in MassBank highlights the need for further support for the development of HRMS libraries. The JBSS samples were screened for the presence of these compounds using DSFP in batch mode. 80 out of the 777 substances were detected. The relatively high detection rate (10.3%) indicates that the screened compounds were of widespread use.

Industrial chemicals such as phthalate esters and phosphates, pharmaceuticals (phenazone and its TPs, carbamazepine, fenbendazole) and pesticides (atrazine, terbutylazine, chloridazon, ametryn, metolachlor, simazine, diuron) appeared to be the most widely detected CECs in the samples. 60 out of 80 detected CECs were monitored by wide-scope target screening and their occurrence, spatial distribution and risk assessment are discussed in detail in the final scientific report of the JBSS 2016 **[132]**. The remaining 20 CECs included industrial chemicals such as phosphates and phthalate esters (triisobutyl phosphate, tris(2-butoxyethyl) phosphate (TBEP), dicyclohexyl phthalate), surfactants (N,N-bis(2-hydroxyethyl)dodecanamide, lauric isopropanolamide), plasticizers and industrial intermediates (benzenesulfonamide, N-butylbenzenesulphonamide) and pharmaceuticals-antimicrobial substances (8-hydroxychinolin).

The occurrence of these 20 compounds is represented in the **SI**, section **S5.7**. Triisobutyl phosphate, with an annual production tonnage of 1,000-10,000 t/a, was detected in all sediment samples and in 45 out of 55 seawater samples, while it was not detected in any biota sample. The spatial distribution of this CEC indicates input from Ukrainian and Georgian shores to the Black Sea (**Figure S5-4F (SI)**). The plasticizer N-butylbenzenesulfonamide, produced in the same tonnage range, was detected in 13 out of 55 seawater and in 17 out of 19 sediment samples. TBEP, produced in the same tonnage range, was detected in 7 out of 55 seawater and 10 out of 19 sediment samples. The surfactant N,N-bis(2-hydroxyethyl)dodecanamide, with the use of 100-1,000 t/a, was detected only sporadically in one seawater, three sediment and three biota samples. Finally, another surfactant lauric isopropanolamide was detected in 32 seawater samples, whereas its presence was not observed in biota or sediment samples.

5.2.9 Conclusion and future developments

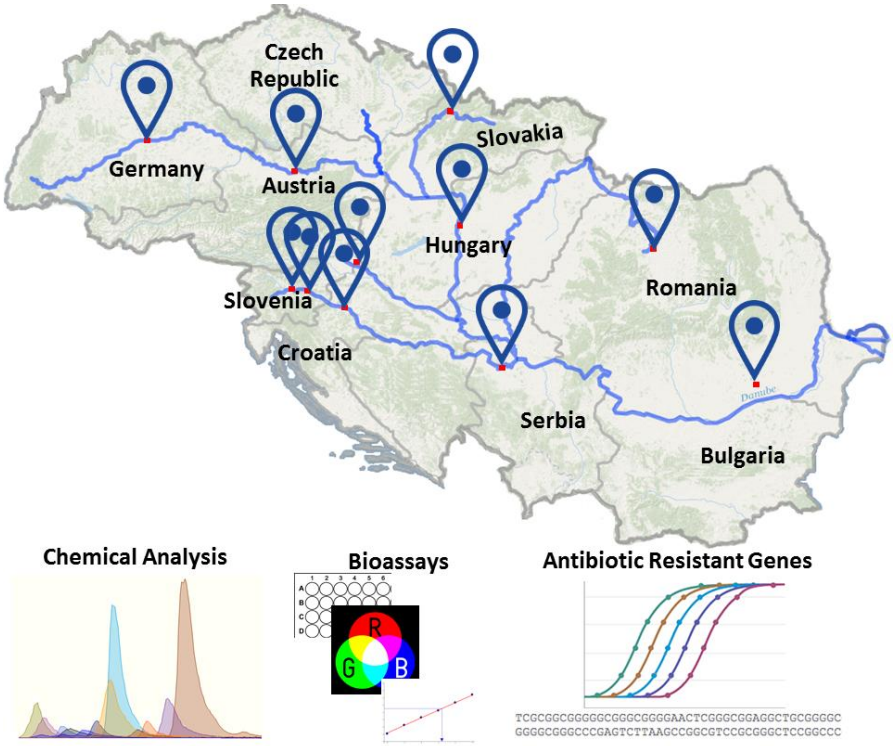
An open, integrated, interactive and intuitive platform for archiving, processing and data mining from a large amount of LC-HRMS data produced by the environmental community was developed and thoroughly tested. The platform allows for the retrospective suspect screening of the presence of tens of thousands of CECs and their TPs in environmental samples in a systematic and consistent way, with a goal of becoming a European and possibly global standard. The platform integrates tools for storing raw HRMS chromatograms of samples, each containing typically several thousands of compounds, in a uniform mzML format independent from vendor software. Both single substance and batch queries are possible across selected or all of the samples stored in the platform.

The results of the NTS workflow used in DSFP were validated against the outcomes of the target screening of 2248 substances in the same samples. The compounds identified by both approaches overlapped in 97% of cases for seawater, 94% for biota and 106% for biota samples. The applicability and the potential use of DSFP was demonstrated in the screening of 670 antibiotics and 777 REACH chemicals in Black Sea seawater, sediment and biota samples. Thus, DSFP incorporates all the state-of-the-art developments of HRMS screening methodologies.

Continuous improvements to the features of the platform and the addition of modules that provide enhanced data processing capabilities remain a priority. Next steps discussed within the NORMAN network are directed towards collecting a critical mass of raw mass chromatograms covering samples from all environmental compartments for quick screening of a presence of major polluting compounds across Europe and beyond. Finally, the integration of GC-APCI-HRMS and GC-EI-HRMS data is under progress as a significant upgrade towards a unified global platform for storing, viewing and screening of much wider analytical window of environmental pollutants.

CHAPTER 6

CHARACTERIZATION OF WASTEWATER EFFLUENTS IN THE DANUBE RIVER BASIN WITH CHEMICAL SCREENING, IN VITRO BIOASSAYS AND ANTIBIOTIC RESISTANT GENES ANALYSIS



Highlights

- 280 out of 2248 target substances were detected
- Detected chemicals were evaluated based on their ecotoxicological properties
- 10 *in vitro* bioassays were applied for assessing the adverse effects in wastewater
- Effect-based risk assessment of WWTP effluents was applied
- 13 antibiotic resistant genes were determined in the Danube River Basin

This case study has been accepted for publication in Environment International

6.1 Introduction

The Danube River Basin (DRB) is the world's most international river basin covering a total area of 801,463 km^2, including territories from 19 countries. DRB is the Europe's second largest river basin and serves more than 80 million people by providing drinking water, industrial and agricultural water supply, hydroelectric power generation, tourism and fisheries among others **[146]**. Therefore, careful management of DRB's water resources is needed, including control over chemical pollution. EU environmental legislation aims to protect all European water bodies by achieving their good chemical and ecological status **[9]**. In the content of chemical status, the EU WFD established a list of 45 priority substances **[9]**, supplemented by a set of additional 15 compounds **[10]** in the recently revised Watch list, which are required to be monitored by Member States and benchmark their concentration against the EQS.

Despite the regulatory efforts, many toxic anthropogenic chemicals are released into the environment that may have an adverse effects on the human health, ecosystem and diminish the quality of the aquatic resources **[12]**. Despite large investments into WWTP technology, point source discharges from sewage plants of big cities in the DRB (e.g., Vienna, Bratislava, Budapest, Belgrade and Bucharest) still represent a significant route of input of numerous contaminants into the river **[147, 148]**. The introduction of untreated or partially treated wastewater generates complex chemical mixtures, which may impact severely the ecosystem of the receiving surface water, as shown recently in the case of Novi Sad **[149]**. To assess such complex chemical mixtures, it is necessary to investigate their overall toxic potency and prioritise frequently occurring compounds based on their ecotoxicity **[149]**.

There are many studies at regional level investigating the occurrence of specific classes of emerging substances in wastewater, such as psychoactive substances **[150]**, benzodiazepines **[151, 152]**, opioid analgesics **[153]**, and perfluorinated substances **[154]**. There are considerably fewer studies focusing on bioassay applications **[7, 155]** and on AR **[156, 157]**, whereas only a very limited number of studies are dealing with combined wide-scope chemical and bioassays screening to assess the quality of wastewater **[7, 149, 155]**. Also, the DRB was assigned as a 'reservoir of AR' in the Joint

Danube Survey 3 **[143]**, one of the most serious threats to human health **[135]**. It has been clearly recognized that more information is needed to define the composition of typical chemical mixtures, their fate and adverse effects on the environment, and to establish a comprehensive risk assessment scheme allowing the regulators to define preventive action plans within the programs of measures **[9]** at a local, national or river basin scale. Here, bioassays covering a range of the ecotoxicity spectrum, as wide as possible, are considered as the key instrument in assessing the mixture toxicity **[158, 159]**.

To facilitate an overview of effluent wastewater released into the Danube River and its tributaries, twelve WWTP effluent samples of various size and using different treatment technologies from nine countries were collected in cooperation with the International Commission for the Protection of the Danube River (ICPDR; 14 European countries and European Commission). The objectives of this study were to: (i) evaluate the occurrence of CECs using the state-of-the art wide-scope chemical screening techniques; (ii) apply NORMAN prioritisation framework **[160, 161]** to prioritise the detected substances; (iii) apply a battery of bioassays to assess the adverse effects of mixtures of pollutants (exceedance of effect-based trigger values associated with various modes of action); (iv) test the feasibility of the newly proposed risk assessment scheme based on bioassays responses, and (v) assess the occurrence of antibiotics and antibiotic resistant genes (A&ARGs) in the collected wastewater effluents.

6.2 Materials and methods

6.2.1. Study area and sampling

The WWTPs in the DRB were selected in a way representing each country's predominant wastewater treatment technology including large plants in country capitals (e.g. Budapest, Ljubljana, Bucharest and Zagreb), large cities (e.g. Brno, Cluj-Napoca, Žilina and Augsburg) and towns (e.g., Amstetten and Varaždin). All plants reported to receive both municipal and industrial wastewater. Wastewater from Augsburg and Vipap consisted mainly of industrial wastewater (65% and 73% respectively). An overview of the sampling stations, their location, their annual average daily wastewater discharge and treatment type can be found in **Table 3**.

Composite seven-day wastewater effluent samples were collected during dry weather and under normal operating conditions. Basic physico-chemical parameters and information about the plants were provided by the WWTPs' operators. Samples for analyses of organic substances remained in the freezer at -20 ^{o}C in the WWTP and frozen during transport. Each WWTP collected 7 L of composite wastewater (1 L for every day for a week). The 7 L samples from each WWTP were mixed to form the weekly averaged sample. 2 L from the weekly averaged samples were processed for chemical analysis, 0.5 L were sent to KWR (Nieuwegein, Netherlands) for analysis of antibiotic resistant genes (ARGs) and 1.0 L were sent to BDS (Amsterdam, Netherlands) for application of bioassays. All samples were processed immediately after arrival to the laboratory.

6.2.2. Chemical analysis

Details for chemicals and reagents used for chemical analysis are given in section **S6-1 (SI)**. Samples were cleaned up and preconcentrated 4000 times on Atlantic HLB-M Disk using HORIZON SPE-DEX 4790 (USA) with 47 mm disk holder according to the extraction program described in section **S6-2 (SI)**. Extracts were evaporated using gentle stream of nitrogen and reconstituted with 500 µl of 50:50 methanol:water for UHPLC-ESI-QTOF and UHPLC-ESI-QqQ analysis. Before instrumental analysis extracts were filtered through RC syringe filters of 4 mm diameter and 0.2 µm pore size (Phenomenex, USA).

UHPLC-ESI-QqQ instrumental analysis was performed for the highly-sensitive determination of 158 pharmaceuticals, drugs of abuse and their TPs **[94, 162]**. A Thermo UHPLC Accela system connected to a TSQ Quantum Access triple quadrupole mass spectrometer from Thermo Electron Corporation (San Jose, CA, USA) equipped with an electrospray ionisation source (Thermo IonMAX) in both positive and negative mode. Chromatographic separation was performed on an Atlantis T3 C18 (100 mm x 2.1 mm, 3 µm) column from Waters (Milford, MS, USA) at a constant flow rate of 100 µL min^{-1}. The mobile phase, the gradient elution program and the ESI parameters are presented in section **S6-3 (SI)** and the optimised ionisation mode, fragmentation

voltages, collision energies and chromatographic retention times for each analyte are summarized in **Table S6-3A (SI)**.

Table 3. Sampling stations, location, coordinates, population equivalent, annual average daily wastewater discharge and treatment type. Stations ordered by the average annual daily wastewater discharge.

Sampling station	Country	Longitude	Latitude	Population Equivalent	Annual average daily wastewater discharge (m^3 day^{-1})	Treatment type
Bucharest	Romania	44°23'38.6"N	26°13'49.3"E	1,327,995	502,789	Tertiary
Zagreb	Croatia	45°47'23.0"N	16°05'14.3"E	842,425	340,000	Secondary
Budapest	Hungary	47°27'23.0"N	19°04'16.5"E	1,174,643	248,904	Tertiary
Augsburg	Germany	48°24'31.4"N	10°53'07.7"E	659,387	138,964	Tertiary
Cluj-Napoca	Romania	46°47'39.2"N	23°40'56.2"E	382,031	115,893	Tertiary
Brno	Czech Republic	49°07'51.1"N	16°37'53.9"E	397,945	93,010	Tertiary
Ljubljana	Slovenia	46°04'15.1"N	14°37'15.2"E	537,712	73,041	Secondary
Žilina	Slovakia	49°14'50.3"N	18°38'45.0"E	139,934	53,570	Tertiary
Varaždin	Croatia	46°18'34.8"N	16°23'43.2"E	97,450	23,443	Secondary
Amstetten	Austria	48°06'31.7"N	14°53'43.2"E	150,000	23,225	Tertiary
Sabac	Serbia	44°44'38.3"N	19°43'37.0"E	84,000	13,833	Tertiary
Vipap (Krško)	Slovenia	45°57'00.3"N	15°29'33.4"E	152,487	13,682	Tertiary

UHPLC-ESI-QTOF analysis was performed using a UHPLC apparatus (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Dreieich, Germany), coupled to the QTOF-MS mass analyzer (Maxis Impact, Bruker Daltonics, Bremen, Germany). Chromatographic separation was performed on an Acclaim RSLC C18 column (2.1 x 100 mm, 2.2 µm) from Thermo Fisher Scientific preceded by a guard column of the same packaging material, kept at 30˚C. Gradient program, ESI parameters and mobile phases are summarized in **Table S6-4 (SI)**. Samples were subjected to wide-scope target screening

of 2248 compounds (list of compounds "UATHTARGETS" **[137]**) according to the established screening method **[144]**.

6.2.3. Bioassays

Detailed sample preparation protocol using fully validated methods and standard operational procedures are described in section **S6-5 (SI)**. The CALUX® bioassays (Chemical Activated Luciferase eXpression; BioDetection Systems BV, Amsterdam, the Netherlands) applied in the present study utilise cell lines, incorporating the firefly luciferase gene coupled to Responsive Elements (REs) as a reporter gene for the presence of compounds able to activate the respective REs. Cell culture information for the applied CALUX® bioassays (cell type, cell type species, % DMSO, fold dilution, % CO_2, exposure time, confluence, medium and additions to medium) can be found on **Table S6-5.1 (SI)**. Cells that were exposed to compounds of interest not only express proteins that are under normal circumstances associated to RE, but also luciferase. By addition of the appropriate substrate for luciferase, light is emitted. The amount of light produced is proportional to the amount of ligand-specific receptor binding, which is benchmarked against the relevant reference compounds (17ß-estradiol, Flutamide, Dexamethasone, Ru486, GW7647, Roziglitazone, B[a]P, Curcumine and Nicardipine for the ERα, anti-AR, GR, anti-PR, PPARa, PPARg, PAH, Nrf2 and PXR CALUX®, respectively). In this way, the CALUX® bioassays report 17ß-estradiol, Flutamide, Dexamethasone, Ru486, GW7647, Roziglitazone, B[a]P, Curcumine and Nicardipine equivalents, respectively. To test for possible cytotoxic effects of the sample extracts, the cytotox CALUX® activity was also determined. Cytotox CALUX® cells constitutively express luciferase. In case cytotox CALUX® cells are exposed to sample extracts causing cytotoxicity, a decrease in luminescence is observed. A reduction of 20% in luminescence is considered as a cytotoxic response. To facilitate water quality assessment, effect-based trigger values (EBTs) have been established. Bioassay responses above EBTs indicated potential risk of adverse effects to the ecosystem. EBTs for the used CALUX® bioassays were retrieved from the literature **[163, 164]** and are presented in **Table S6-5.2 (SI)**. The signal from each of

the bioassays was compared with the EBT and thus allowed for ranking of the toxicity of wastewater effluents.

6.2.4. Multiplex qPCR Assays

Aliquots of 100-250 mL were filtered through a 0.22 µM CA (Cellulose Acetate) pore filter membrane (Sartorius, Göttingen, Germany), depending on the concentration of particles in the sample. DNA extraction was performed using the DNeasy PowerSoil Kit (QIAGEN Benelux B.V), following the manufacturer's instructions with two exceptions: (i) a filter was used as material instead of soil and (ii) 10 µL/sample of internal control (IC) at a concentration of 2.5E+04 gene copies/µL were added to buffer 2 in order to be able to quantify DNA loss during the extraction process and detect potential qPCR reaction inhibition. DNA extracts were stored at -20 °C prior to qPCR analysis. qPCR analysis was performed within 2 weeks of DNA extraction. qPCR assays were performed for a total of 13 ARGs, for one mobile genetic element (*intI*1) as proxy of the anthropogenic pollution **[165]** and the IC using multiplex qPCR assays. 16S rRNA was quantified using a SYBR Green qPCR assay. The following ARGs were quantified by qPCR: *aph*(III)a, bla_{KPC}, bla_{OXA}, bla_{SHV}, *erm*B, *erm*F, *mec*A, *qnr*S, *sul*1, *tet*B, *tet*M, *van*A and *van*B. Primer and probe sequences and conditions of multiplex qPCR assays are described in section **S6-6 (SI)**. 10 µL of DNA extract, of positive control or of negative control were used per well for qPCR assays. Each qPCR assay was performed with undiluted DNA extract, in an initial first qPCR run, and with 1:10 diluted DNA extract, in subsequent runs. This was done to detect potential inhibitors and inhibition of the qPCR reactions for each sample. A positive control and a negative control were included in every assay to ensure multiplex qPCR integrity. Multiplex standards consisted of gBlock fragments containing relevant gene sequences for the three genes within a multiplex qPCR assay. The 16S rRNA standard consisted of a plasmid containing the relevant 16S rRNA gene sequence. In both cases, standards were made up of 5 subsequent dilutions with concentrations ranging from 2.5E+04 to 2.5E+00 gene copies/µL. Multiplex qPCR assays were performed using the iQ™ Multiplex Powermix (Bio Rad, Munich, Germany) and the qPCR reaction was performed using a CFX96™ Real-Time PCR Detection System (Bio Rad, Munich,

Germany). CFX96™ Real-Time PCR Detection System were interpreted by CFX Manager v.3.1.1517.0823. Multiplex qPCR data analysis is described in section **S6.6.2 (SI)**.

6.2.5. Quality assurance and quality control

The chemical method used in the present work was evaluated in terms of linearity, accuracy, sensitivity, repeatability and matrix effects. Seven-point calibration curves (0.5–100 ng mL^{-1}) were generated using linear regression analysis. The linearity was qualified by linear correlation coefficient (r^2). Accuracy of the method was assessed with recovery experiments in effluent wastewater samples at two concentration levels (10.0 and 100.0 ng L^{-1}). Extraction recovery was calculated by dividing the peak area of the spiked samples by the peak area of the matrix-matched samples (extracts spiked at the end of the sample preparation). As real samples may already contain target compounds, wastewater samples were analyzed to determine their concentrations, which afterwards were subtracted from the spiked and the matrix-matched samples. Method repeatability was evaluated with calculation of intermediate precision and was expressed in terms of relative standard deviation (% RSD) at the same concentration levels (10.0 and 100.0 ng L^{-1}). Matrix effect was expressed as percentage of suppression or enhancement was calculated using the following equation: %Matrix Effect = (Matrix Factor - 1) × 100, in which matrix factor was the fraction of the peak area of the matrix-matched samples divided by the peak area of the standard solutions. More details about quality assurance and quality control can be found in the **section S6.7.1 (SI)**. All samples were spiked for 31 internal standards (**Table S6-1**). Quantification was based on standard additions, and isotopically labelled compounds were used only for the quantification or those compounds in which isotopically analogues compounds were available. A field blank and a laboratory procedural blank were used to detect any unwanted contamination. The blank samples accompanied the wastewater samples for all types of analysis (chemical, bioassays and ARGs analysis). The signals observed in blank samples were subtracted. Octocrylene was the only case in which the signal of the blank samples exceeded the signal in the wastewater sample and thus was excluded from the results.

For bioassays testing wastewater samples (500 mL) were extracted by means of Solid Phase Extraction (SPE) according to the fully-validated BDS protocol (p-bds-096). To test for possible cytotoxic effects of the samples analysed, the cytotox CALUX activity was determined. For the determination of the various CALUX activities, CALUX cells were seeded in 96 wells plates in assay medium. Following exposure of the CALUX cells to serial dilutions of the sample extracts in triplicate, the induction of luciferase production was quantified by measuring luminescence following addition of the substrate luciferin. On each 96-well plate, a complete calibration curve for each respective bioassays is also analysed using the relevant reference compounds. Analysis result of the test samples are intrapolated in the calibration curve for quantitative determination of (ant)agonistic potential of the test samples. Only dilutions that did not show any signs of cytotoxicity (relative induction in the cytotox CALUX bioassay > 80%) were used for final evaluation of CALUX analysis results. The bioassays were performed according to standard BDS standard operating procedures and protocols for culturing U2OS CALUX cells (p-bds-083), for analysis of luciferase activity in the PAH CALUX bioassay (p-bds-066), for analysing samples with U2OS CALUX bioassays using sigmoidal dose response curves (with 0.1% or 1% DMSO; p-bds-085), for harvesting the cells and measurement (p-bds-070), and for calculating U2OS CALUX results using sigmoidal dose response curves (p-bds-084).

To assess ARG extraction 10 µL of internal control (IC) at a concentration of 2.5E+04 gene copies μL^{-1} were added in order to quantify DNA loss during the extraction process and detect potential qPCR reaction inhibition. All samples were processed within 12 hours of their arrival. DNA extracts were stored at -20 °C prior to qPCR analysis. qPCR analysis was performed within 2 weeks of DNA extraction. All qPCR assays were performed in triplicates. Each qPCR assay was performed with undiluted DNA extract, in an initial first qPCR run, and with 1:10 diluted DNA extract, in subsequent runs. This was done to detect potential inhibitors and inhibition of the qPCR reactions for each sample. A positive control and a negative control were included in every assay to ensure multiplex qPCR quality.

6.2.6. Prioritisation of chemicals

Risk assessment of the detected target compounds was based on the prioritisation methodology developed by the NORMAN network **[160, 161]**. The method is based primarily on comparing the measured concentrations of detected substances against their Provisional No Effect Concentration (PNEC), which represent their ecotoxicological threshold values. In cases when no experimental data on the toxicity of detected substances were available, predicted PNECs (P-PNECs) were derived by QSTR models **[34]**. All PNEC values used in this study were extracted from the NORMAN ECOTOX database (https://www.norman-network.com/nds/ecotox/). Risk score was calculated as the sum of three indicators: (i) Frequency of Appearance (FoA); (ii) Frequency of PNEC exceedance (FoE), and (iii) extent of PNEC exceedance (EoE). The first indicator indicates in how many sites the compound was detected above the limit of detection (LOD). The second indicator considers the frequency of monitoring sites with observations of a compound above a certain effect threshold. For the calculation of this indicator, a compound's maximum observed concentration at each site (MECsite) is compared to the lowest PNEC. Subsequently, the number of sites where the threshold was exceeded was divided by the total number of sites where the respective compound was monitored. The third indicator ranks compounds with regard to the extent of the effects expected. It is defined as the 95th percentile of all MECsite values per compound (MEC95) divided to the PNEC. The resulting hazard ration was then scaled from 0 to 1. The Risk Score is the linear combination of the indicators scaled from 0 to 1. In the end, only compounds with a priority ranking value of more than 1.01 were listed. For the remaining substances, risk was assumed to be negligible. More details about the prioritization scheme used can be found in the study of Slobodnik et al. **[167]**.

6.3 Results and discussion

6.3.1. Occurrence and spatial distribution of chemicals

Compound names, category, molecular formula and SMILES chemical identifier for all (281) compounds detected in at least one sample are summarised in **Table S6-7.2A (SI)**. The detected CECs were grouped in seven categories: Pharmaceuticals (101), Pesticides (42), Psychoactive drugs (40), Industrial chemicals (34), Antibiotics (32) and Drugs of abuse, tobacco ingredients and steroids (26), artificial sweeteners (6). The sum of the concentration of all compounds (indicated as "cumulative concentration") corresponded to 6,615-27,003 ng L^{-1} depending on the sampling location (**Table S6-7.2B (SI)**). Effluent wastewater from large WWTPs of capital cities showed the highest cumulative concentration; Ljubljana (27,003 ng L^{-1}), Bucharest (21,234 ng L^{-1}), Budapest (20,257 ng L^{-1}), Zagreb (17,004 ng L^{-1}). The lowest cumulative concentration was observed for Varazdin (7,408 ng L^{-1}) and Augsburg (6,615 ng L^{-1}).

Pharmaceuticals were not only the most often detected (101), but also the most ubiquitous class of substances (in terms of concentration) in all samples (**Figure 12**). They represented 25-67% of the total concentration of the target substances. The highest concentration for pharmaceuticals was observed for plants serving the capital cities (53-67%). Industrial chemicals (5-30%) and pesticides (3-21%) proved to be the second and third most abundant compound classes. In cases in which concentrations of pharmaceuticals did not exceed 50% of the total concentration (Augsburg, Amstetten, Zilina, Varazdin and Brno), an elevated concentration were observed for pesticides and industrial chemicals. Psychoactive drugs (3-23%) and antibiotics (2-17%) proved to occur in lower concentration than industrial chemicals and pesticides. The highest antibiotic composition was observed for Ljubljana and Sabac (17 and 16% respectively). Despite of its size, Sabac showed elevated concentrations of antibiotics and ARGs, which can be partially attributed to the pharmaceutical industry (Sabac; production of erythromycin, sulfamethoxazole and ciprofloxacin). Drugs of abuse (2-9%) showed similar occurrence levels in all samples. A detailed description of the occurrence of the individual substances and their TPs is discussed in detail at section **S6.8 (SI)**.

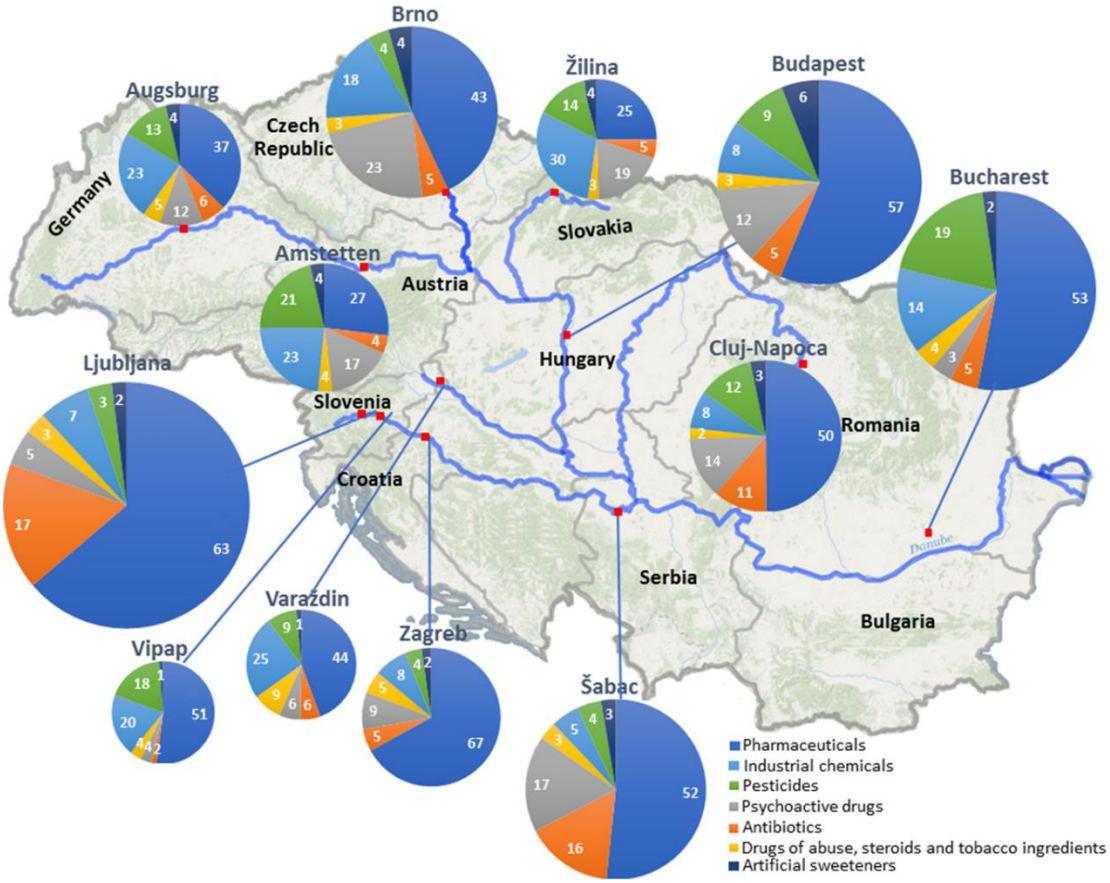


Figure 12. Composition of chemicals in the effluent wastewater samples collected in the Danube River Basin. Pie plots for each sampling station (marked in red) represent the percentage (%) of each class of target compounds, whereas size of the pie plots is proportional to the cumulative concentration of all detected target compounds.

6.3.2. Occurrence of ARGs in wastewater effluents

11 out of the 13 genes, and *int1* were detected in at least one sample (Section **S6-1,** SI). Six ARGs (Aph(III)a, bla_{OXA}, *erm*B, *erm*F, *sul*1 and *Ttet*M), and *int*1, seemed to have wide-spread occurrence, since they were detected in all samples. Five ARGs (bla_{SHV}, *mec*A, *qnr*S, *tet*B and *van*A) were detected sporadically, while bla_{KPC} and *van*B remained undetectable (**Figure 13 and Figure S6-9A (SI)**). Relative concentration levels of *erm*B, *sul*1 and *tet*M seemed to be constant in all the investigated samples (Welch's ANOVA, p>0.05), with concentrations ranging from 1.8E-05 to 4.9E-03 gene copies normalized to 16S rRNA. *Aph*(III)a, *intI*1, *erm*F and bla_{OXA} fluctuated widely with relative concentrations ranging from 2.4E-08 to 3.1E-02 gene copies per 16S rRNA

copy, respectively. *Van*A was detected in two sampling stations (Varazdin and Bucharest). *Mec*A, *qnr*S and *tet*B were detected in relative concentrations ranging from 2.39E-08 to 2.35E-04 gene copies per 16S rRNA.

The most polluted sampling locations by ARG presence and abundance were Varazdin, Bucharest and Sabac. Varazdin and Bucharest were the locations at which 12 out of 14 ARGs were detected. High concentration of ARGs in Varazdin can be attributed to the extensive agriculture (e.g. poultry farming) in the region. The minimum number of detected ARGs was observed at the WWTPs Amstetten and Augsburg (seven genes). WWTPs Bucharest and Varazdin were also the sampling stations with the highest cumulative ARG concentration in gene copies/mL, whereas WWTPs Amstetten and Augsburg were the locations with the lowest cumulative ARG concentration (**Figure 13**). The highest absolute concentrations (gene copies/mL) were observed for four ARGs (*Aph*(III)a, *tet*M, *van*A, *mec*A) at the WWTP Bucharest, four ARGs (bla_{OXA}, bla_{SHV}, *qnr*S, *tet*B) at the WWTP Sabac, two ARGs (*erm*B, *sul*1) at the WWTP Varazdin and one ARG (ermF) at the WWTP Brno. Three of the detected ARGs (bla_{SHV}, *tet*B and *van*A) were found in 50% or less samples. From the investigated ARGs, *intI*1 and *sul*1 were the most abundant in absolute and relative concentration. *IntI*1 has previously been suggested as an indicator for ARG pollution **[165]**. The present findings further affirm this suggestion, as the three most polluted sampling location coincided with the highest measured *intI*1 concentrations. In most of the cases, the concentration of the antibiotics and the ARGs did not correlate. Exception to this trend was *qnr*S, which correlated significantly with quinolones, having a correlation coefficient of 0.77 (Hollander and Wolfe test, p-value=0.009). It has previously been shown that elevated ARG concentrations from point sources, such as WWTPs, can have a significant and lasting impact on downstream water bodies **[168-171]**, which makes information on ARG concentration in WWTP effluent necessary for risk assessment. WWTP effluents are of special interest as treatment has shown the potential to increase the frequency of antibiotic resistant bacteria **[172]**.

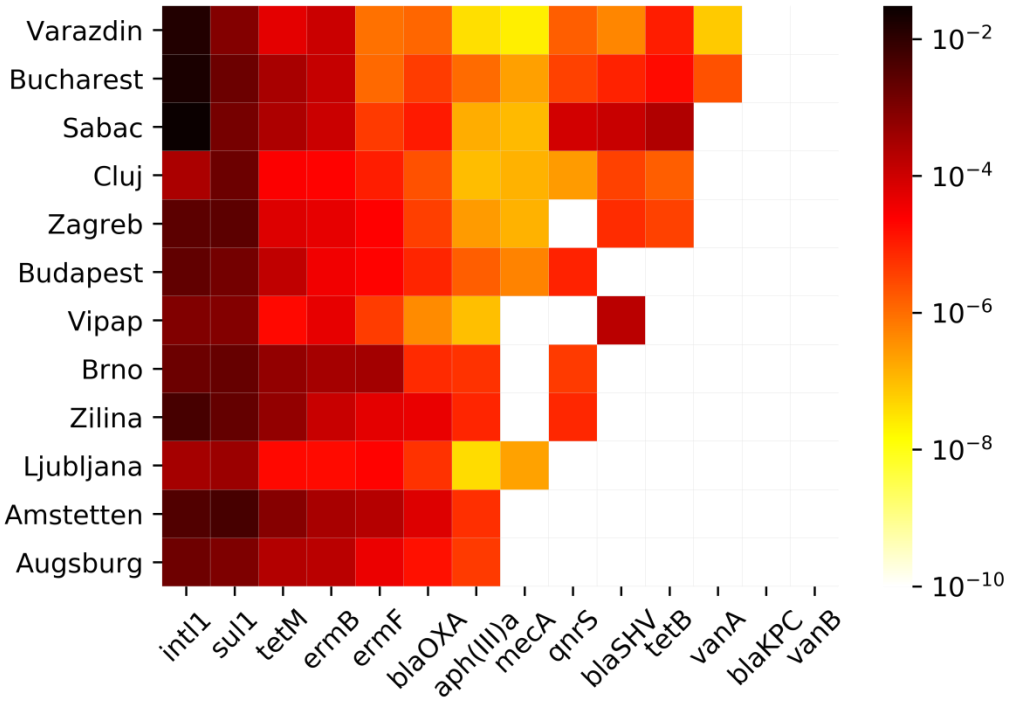


Figure 13. Gene presence and concentration across the different sampling sites (WWTPs) in the Danube River Basin. Shown are relative gene copies (normalized to 16S rRNA).

6.3.3. Risk assessment

Table 4 shows the 17 out of the 281 compounds which were prioritised. PFOS was prioritised first, exceeding the established EQS PNEC in all wastewater samples. The compound received the attention of other researchers in the past, who reported the compound in concentrations higher that EQS limit set by WFD (1 ng L^{-1}); medium concentration 7 ng L^{-1} in Danube JDS2 **[173]** samples, 5.9 ng L^{-1} in Danube JDS3 samples **[174]**, up to 33 ng L^{-1} in wastewater and river water from Slovenia **[154]** and 95th percentile measured experimental concentration (MEC_{95}) 31 ng L^{-1} in four European catchments **[175]**. The second prioritised compound was the antibiotic ofloxacin with a risk score 2.58, followed by telmisartan with a risk score 2.57. There are not many reports for the occurrence of ofloxacin and telmisartan in wastewater from the DRB region. However, previous reports for ofloxacin in Europe showed ecotoxicological important concentration levels (up to 507 ng L^{-1} in Spain **[176]** and

55.5 ng L^{-1} for United Kingdom **[177]**. Ofloxacin requires the attention of regulators and researchers for further monitoring of the compound in the catchment and conclusion whether it should be included in the legislation. Telmisartan exceeded P-PNEC value for 83 % of the samples and has not been monitored adequately in the DRB (only a report in Hungarian wastewater at concentration up to 4800 ng L^{-1} **[178]**). It worth noticing that telmisartan reported to exceed P-PNEC in seawater samples from Black Sea in Joint Black Sea Surveys (JBSS; EU/UNDP EMBLAS-II project; 2016). Further ecotoxicological experiments are required to verify P-PNEC value, so that robust conclusions are reached whether occurrence of telmisartan in the samples is of concern for the ecosystem. Other similar cases where P-PNEC verification by experimental data is needed to assess the hazard of compounds were the surfactant C12-LAS, the NSAID meclofenamic acid, the insecticide fipronil**,** the metabolite of methadone (EDDP) and of omeprazole (4-hydroxy-omeprazole). C12-LAS was detected above P-PNEC for 83% of the samples and its exceedance may be alarming for rest LAS surfactants, which even though detected in the samples could not be reliably quantified, because of lack of standards. Meclofenamic acid exceeded P-PNEC at 67% of the samples, whereas fipronil at 58% of the samples. Fipronil was reported by previous studies at lower concentration levels (0.4-4.5 ng L^{-1} in Danube in Novi Sad **[149]** and MEC_{95} 1.51 ng L^{-1} in EU catchments **[175]**). The high concentration reported here can be attributed to the fact that collected samples were wastewater and not surface water, which indicates high proximity to the source of pollution. Fipronil needs to be monitored in more samples to reach safe conclusions. Same observations were valid for carbendazim**,** which exceeded PNEC for 67 % of the samples, but it was previously reported at lower concentration levels in literature (4.0-12.4 ng L^{-1} in wastewater from Novi Sad **[149]** and MEC_{95} 2.8 ng L^{-1} in EU catchments **[175]**). Carbendazim and fipronil screening in more samples and in surface water samples is further needed to conclude whether their occurrence induces hazards to ecosystem.

Other prioritised substances were the NSAIDs diclofenac and ibuprofen, carbamazepine, 4-tert-octylphenol and venlafaxine. All these substances are known to be widespread in the environment and are well-studied (EQS PNEC available and occurrence levels) in the DRB. Their concentration exceeded EQS PNEC for 92%, 83%,

67%, 58% and 50% of the samples for diclofenac, carbamazepine, 4-tert-octylphenol, venlafaxine and ibuprofen respectively. The NSAIDs diclofenac and ibuprofen have been reported to exceed PNEC in several studies **[147, 174, 179, 180].** Both compounds are suggested for regular monitoring in the DRB catchment. Carbamazepine has been detected at concentrations levels 120-1550 ng L^{-1} in wastewater from Balkan region **[147]**, in wastewater from Romania at concentration ranges 213-774 ng L^{-1} **[181]** and in surface water of Danube river (average concentration of 33 ng L^{-1} for JDS2 samples **[173]** and 25 ng L^{-1} for JDS3 samples **[174]**). Average concentration in diluted matrices such as Danube water may dropped below PNEC. However, the introduction of large quantity of this drug in the ecosystem may have negative effects at least in regional level.

Table 4. Compounds prioritised based on the risk score, which is the sum of Frequency of Appearance (FoA), Frequency of Exceedance (FoE) and Extent of Exceedance (EoE). Table presents only compounds with risk score more than one.

Name	**LOQ [µg/L]**	**PNEC [µg/L]**	**PNEC type**	**MEC [µg/L]**	**FoA**	**FoE**	**EoE**	**Risk Score**
PFOS	0.0003	0.001	EQS WFD	0.05	1.00	1.00	1.00	3.00
Ofloxacin	0.005	0.021	P-PNEC exp. Aquire 80421	3.1	1.00	0.75	0.83	2.58
Telmisartan	0.003	0.042	P-PNEC	2.3	1.00	0.83	0.74	2.57
Diclofenac	0.005	0.050	EQS-proposal	1.4	1.00	0.92	0.45	2.37
Dodecyl-benzenesulfonate (C12-LAS)	0.010	0.086	P-PNEC	1.8	1.00	0.83	0.24	2.07
Carbamazepine	0.0002	0.050	PNEC chronic Aquire 152195	0.7	1.00	0.83	0.17	2.01
Ibuprofen	0.001	0.010	EQS	1.0	0.50	0.50	0.92	1.92
4-tert-Octylphenol	0.010	0.100	EQS	0.3	1.00	0.67	0.06	1.72
Meclofenamic Acid	0.009	0.097	P-PNEC	0.3	1.00	0.67	0.05	1.72
Fipronil	0.003	0.023	P-PNEC	0.4	1.00	0.58	0.13	1.71
Carbendazim	0.026	0.150	AA-QS water eco INERIS (2017)	1.1	0.92	0.67	0.10	1.68
Venlafaxine	0.006	0.038	EQS-proposal	0.1	0.92	0.58	0.05	1.55
Clarithromycin	0.001	0.120	EQS-proposal	0.7	1.00	0.42	0.05	1.47
4-Hydroxy-Omeprazole	0.002	0.263	P-PNEC	8.5	0.92	0.08	0.13	1.13
EDDP	0.0001	0.137	P-PNEC	0.2	0.83	0.25	0.02	1.10
Temazepam	0.004	0.071	PNEC chronic Aquire 175030	0.2	0.83	0.17	0.02	1.02
Sertraline	0.0003	0.091	PNEC exp. Aquire 107936	0.1	0.92	0.08	0.01	1.01

Venlafaxine and 4-tert-octylphenol were detected in similar concentration as in previous studies (e.g. venlafaxine previously detected as high as 259 ng L^{-1} **[150]** and up to 272 ng L^{-1} for 4-tert-octylphenol **[147]**, whereas clarithromycin measured at higher concentration when comparing to other studies (e.g. as high as 59.7 ng L^{-1} **[149, 175]**) Last compounds to be prioritised because of the low FoE and EoE scores were temazepam and sertraline, which exceed PNEC for 17 and 8% of the samples respectively.

6.3.4. Application of *in vitro* bioassays

The results acquired by application of CALUX® bioassays expressed per liter of water can be found in section **S6-10 (SI)**. To facilitate the visualization and discussion, results were expressed as fold-induction relative to the LOQ of the respective CALUX® analysis (**Table 5**). Results below LOQ were assigned a fold-induction of 0.5. This visualization allows the comparison of the responses of each CALUX® bioassay among sample locations, but also between different CALUX® bioassays. All samples were proved to be positive for PAH activity and for xenobiotic metabolism (PXR). Highest signal for both PAH and PXR activity was observed for the industrial plant Vipap in Krsko.

The next most frequently detected effect was Estrogenic activity (ERα), with Cluj being the only sample that ERα was not detected. Ljubljana, Bucharest and Varazdin exhibited the highest ERα response than the other samples. Oxidative stress (Nrf2) was detected in 83 % of the samples (not detected in Cluj and Zagreb) at similar equivalents in the samples. Anti-androgenic (anti-AR) and anti-progestin activity (anti-PR) proved to be effects with medium to high FoA, whereas medium detection occurred for glucocorticoid activity (GR). Peroxisome proliferators (PPARα and PPARγ) were scarcely detected.

Table 5. Heat map of CALUX® analysis results for the various WWTPs effluent water sample sites along the Danube river. Values represent the fold-induction of each analysis relative to its respective LOQ. For analysis results that are below the LOQ, the result is represented as 0.5 times the LOQ (0.5). Low activity is marked with green and high activity with red.

	Cytotox CALUX®	anti-AR CALUX®	ERα CALUX®	GR CALUX®	anti-PR CALUX®	PPARα CALUX®	PPARγ CALUX®	PAH CALUX®	PXR CALUX®	Nrf2 CALUX®
Varazdin	4.5	0.5	49	0.5	4.9	2.3	0.6	40	2.3	1.2
Amstetten	0.5	1.9	10	0.5	5.3	0.5	0.5	68	3.6	1.4
Cluj	2.3	2.7	0.5	0.7	12	0.5	0.5	28	2.3	0.5
Augsburg	0.5	0.9	8.8	1.5	1.7	0.5	0.5	38	3.6	1.4
Vipap	0.8	2.7	5.6	0.5	7.1	0.5	0.5	159	9.2	2.1
Budapest	0.5	1.0	5.3	0.5	3.1	0.5	0.5	46	3.0	1.5
Ljubljana	0.5	0.7	60	3.8	3.6	0.5	0.5	17	2.7	1.5
Bucharest	2.0	0.5	69	1.3	6.6	0.5	0.5	22	2.9	3.9
Zilina	0.5	0.8	20	1.0	0.5	0.5	0.5	57	1.1	1.8
Sabac	0.5	1.2	9.5	0.5	0.5	0.5	0.5	57	0.8	1.4
Brno	0.5	1.1	10	1.1	0.9	1.2	0.5	80	1.7	2.4
Zagreb	0.5	0.5	15	0.5	0.5	0.5	0.5	34	1.6	0.5

As happens in case of detected chemicals, not all the detected effects in the samples are harmful to the environment. This happens because bioassays have become sensitive with low limits of detection. The solution is the application of EBTs on CALUX® bioanalysis results as determined for the WWTPs effluent water samples along the Danube River which resulted in the heat map presented in EBT values are as critical for assessing the importance of the observed effects as PNEC values for assessing the ecotoxicity of detected chemicals. Thus, the establishment of robust and reliable EBT values is of crucial importance, because large variations in proposed EBT values may result in misleading conclusions. In context of the presented study, EBTs of PAH CALUX® (6.2 **[164]** and 150 ng B[a]P eq./l sample **[163]**) and PXR CALUX® (54 **[164]** and 3 µg Nicardipine eq./l sample **[163]**) were considered to deviate enough to prevent consolidated conclusions for the importance of the observed effects of the aforementioned bioassays. On the contrary, close EBTs proposed for ERα, Nrf2 and anti-PR CALUX® led to consolidated conclusions. Responses by ERα, oxidative stress and anti-PR CALUX® exceeded the EBT for 92%, 83% and 75% respectively of the investigated samples, while anti-AR® CALUX exceeded the EBT for <25 % of the samples.

Table 6. Application of a typical response plan published effect-based trigger values on CALUX® analysis results of WWTPs effluent water sample from various sites along the Danube river. Colour-coding indicates bioactivities above or below the published EBTs < trigger values > trigger values. a: no trigger value available; b: LOQ of bioassay exceeding EBT.

Sampling stations	Escher et al. (2018)								van der Oost et al. (2017)							
	ERα	anti-AR	GR	anti-PR	PPARγ	PAH	Nrf2	PXR	ERα	anti-AR	GR	anti-PR	PPARγ	PAH	Nrf2	PXR
Varazdin			a		a							a				
Amstetten			a		a							a	b			
Cluj			a		a		b					a	b		b	
Augsburg			a		a							a	b			
Vipap			a		a							a	b			
Budapest			a		a							a	b			
Ljubljana			a		a							a	b			
Bucharest			a		a							a	b			
Zilina			a		a							a	b			
Sabac			a		a							a	b			
Brno			a		a							a	b			
Zagreb			a		a		b					a	b		b	

6.3.5. Putative action plan based on *in vitro* bioassays results

Based on the exceedance of the EBT values, a different response plan for WWTP operators was developed. The sample location and frequency for these bioassays should be linked to specified monitoring requirements for such indications for CECs in these water treatment plant effluents (e.g. frequency of six months; collected at the point of compliance). An exceedance of the above proposed trigger values could initiate the following actions:

(i) If the result is below EBT or LOQ of bioassay (White): No further action required

(ii) If the result is 1-times <EBT< 3-times (Blue): Quality check of data, continue to **monitor every three months**, until 1 year and the EBT < 1

(iii) If the result is 3-times <EBT< 10-times (Green): Data check, immediate re-sampling and analysis to confirm EBT exceedance. **It is also required to quantify specific target compounds which are known to cause the effects observed in the respective bioassay**. Continue to monitor every three months, until 1 year and the EBT< 1

(iv) If the result is 10-times <EBT< 100-times (Orange): All the above actions and enhance **source identification program**. Also monitoring in the distribution system closer to the point of exposure to confirm attenuation of CEC is occurring and to confirm the magnitude of assumed **safety factors associated with removal efficiency, dilution and post-treatment**.

(v) If the results is EBT>100-times (Red): All the above actions. Immediately **confer with the local environmental authorities** to determine the required response action. Confirm plant corrective actions through additional monitoring that indicates the CEC levels are below at least an EBT of 100.

The detailed action category table with EBT values for the applied bioassays can be found in section **S6-11 (SI)**. Application of the described action plan to the WWTP samples collected from Danube catchment resulted in **Table 7**.

Table 7. Typical application of such a response plan of actions for operators of such WWTPs on CALUX® analysis results of effluent-water sample from various sites along the Danube River.

	ERα CALUX®	anti-AR CALUX®	GR CALUX®	PPARγ CALUX®	PAH CALUX®	Nrf2 CALUX®
Varazdin	5	5.7	<19	640	72	51
Amstetten	1.1	22	<20	<520	122	57
Cluj	<0.06	31	34	<420	52	<63
Augsburg	1	10	72	<410	72	57
Vipap	0.65	32	<25	<460	242	92
Budapest	0.56	11	<23	<430	62	58
Ljubljana	6.6	8.4	120	<350	62	62
Bucharest	7.4	5.7	38	<340	82	162
Zilina	2.2	8.9	78	<480	72	75
Sabac	1.1	14	<41	<490	72	57
Brno	0.54	13	47	<1100	122	100
Zagreb	0.8	6	<42	<1100	52	<21

Anti-AR CALUX®: µg Flutamide eq/L; ERα CALUX®: ng 17β Estradiol eq/L; GR CALUX®: ng Dexamethasone eq/L; PPARγ CALUX®: ng Rosiglitazone eq/L; PAH CALUX®: ng Benzo[a]pyrene eq/L; Nrf2 CALUX®: µg Curcumin eq/L

The ERα CALUX® activity observed in in three WWTPs (Varazdin, Ljubljana and Bucharest) would lead to data check, immediate re-sampling to confirm EBT exceedance and chemical analysis of known estrogenicity drivers and to the distribution system to verify attenuation of the drivers. Same response would be in case of PAH CALUX® for ten WWTPs (all wastewater samples with the exception of Cluj and Zagreb). Lower response action (re-sampling, re-analysis to confirm EBT exceedance, chemical analysis of drivers) would be required for ERα CALUX® in four WWTPs (Amstetten, Augsburg, Zilina and Sabac), anti-AR CALUX® in seven WWTPs

(Amstetten, Cluj, Augsburg, Vipap, Budapest, Sabac, Brno), PAH CALUX® in two WWTPs (Cluj, Zagreb) and Nrf2 CALUX® in two WWTPs (Bucharest and Brno). Finally, quality check of data and continuation of bioassay monitoring on a regular basis (every three months) for a year would be required for ERα in four WWTPs (Vipap, Budapest, Brno and Zagreb). Same action would be required for anti-AR in 5 WWTPs (Varazdin, Ljubljana, Bucharest, Zilina and Zagreb), for GR in five WWTPs (Cluj, Augsburg, Ljubljana, Bucharest, Zilina and Brno), for PPARγ in one WWTP (Varazdin) and for Nrf2 in eight WWTPs (Varazdin, Amstetten, Augsburg, Vipap, Budapest, Ljubljana, Zilina and Sabac).

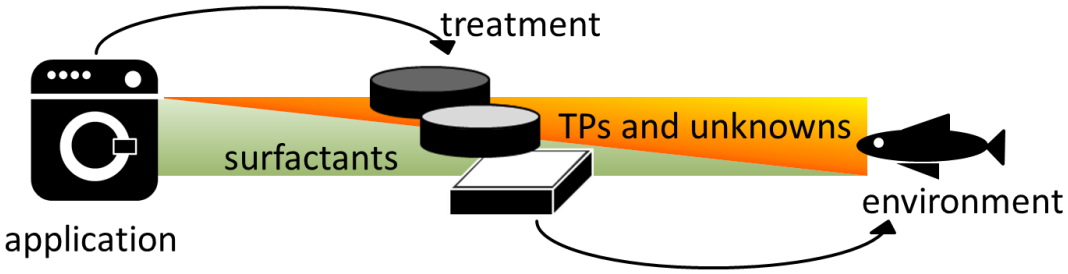
6.4 Conclusions

Representative effluent wastewater samples were collected from nine countries of DRB. The samples were analyzed with the aim to get a holistic overview of the occurrence of chemicals in effluent wastewater using wide-scope target screening methods by LC-QTOFMS and LC-MS/MS. 280 compounds were detected and were subjected in ecotoxicological risk assessment to rank them based on their potential ecotoxicity. 17 out of the 280 compounds were prioritized. The occurrence of PFOS, diclofenac, carbamazepine, ofloxacin and ibuprofen proved to be of concern. Concentration of telmisartan, C12-LAS, meclofenamic acid, EDDP and 4-hydroxy-omeprazole exceeded P-PNEC, but experimental verification of P-PNEC is needed. More occurrence data points were needed for carbendazim and finopril to verify their occurrence and concentration levels. Moreover, the samples were analyzed for a battery of twelve CALUX® bioassays to investigate the effects that chemicals trigger. For this purpose, a set of CALUX® bioassays with a wide-range of mode of actions and established EBT threshold values were selected. The signals obtained by the bioassays were benchmarked against their EBT. Cases with exceedance of EBT were prioritized and a putative action plan was proposed based on the extent of exceedance. The proposed action plan translates the signals from CALUX® bioassays to actions for the WWTP operators. The study highlighted the need for commonly-agreed EBT values, which are needed for the correct translation of the signals from bioassays. Moreover,

the lack of relative effect potency (REP) values for the detected chemicals prevent the connection between chemicals and bioassays and their establishment is crucial towards a better understanding of the pollution. Finally, antibiotic prevalence and abundance of 13 ARGs and one antibiotic resistance mobile element were assessed in the collected samples. Correlation of the concentration of ARGs with antibiotics was investigated. In most of the cases, the concentration of the antibiotics and the ARGs did not correlate with the exception of *qnr*S, which correlated significantly with quinolones. All data collected from these various types of analysis contribute towards a better understanding of the environmental problems caused by organic micropollutants.

CHAPTER 7

OCCURRENCE OF SURFACTANTS AND THEIR TRANSFORMATION PRODUCTS IN WASTEWATER FROM GERMANY



Highlights

- Average concentration was 14.4 µg/L for LAS, 7.4 µg/L for PEGs and 0.6 µg/L for AES
- 1564 surfactants were monitored by target and suspect screening
- High removal rates proved for LAS and AES (99.2% and 99.8% respectively)
- The concentration of LAS-related by-products and TPs surpassed the LAS homologs
- All surfactants and their TPs accounted for up to 82 µg/L in effluent wastewater

This case study has been submitted for publication in Science of the Total Environment

7.1 Introduction

Synthetic surfactants comprise a heterogeneous group of organic compounds that are globally used in large quantities as active ingredients of household and industrial detergents. Due to their surface-active properties, they find broad application in the production of pharmaceuticals and personal care products, paints and varnishes, foodstuffs, plastics, and pesticides **[182]**. Moreover, they are increasingly used in high-technology sectors such as biotechnology and microelectronics in the last decades **[183]**. Surfactants are amphiphilic compounds with a hydrophilic (polar) head and a hydrophobic (nonpolar) hydrocarbon tail, which makes them soluble in polar and nonpolar liquids. They can be classified according to the ionic charge of the hydrophilic part of the molecule (nonionic, anionic, cationic, amphoteric) in the aqueous solution, with anionic and nonionic surfactants accounting for the highest production volumes. According to the European Committee of Organic Surfactants and their Intermediates (CESIO), approximately three million tons of surfactants were manufactured in Western Europe in 2016, about 2.5 times more than 20 years earlier in 1996 **[184]**. Global surfactant production reached 17.6 million tons in 2015 **[185]**.

Linear alkylbenzene sulfonates (LAS) were introduced in 1964 as the readily biodegradable replacements for the commonly used branched alkylbenzene sulfonates (ABS). The substitution of ABS by LAS was stipulated by the excessive foaming of ABS in sewage treatment plants and receiving waters. Today, LAS are one of the most widely used anionic surfactants in detergents, such as laundry powders and liquids, with a proportion of up to 25 percent in consumer products and up to 30 percent in products for professional use **[186]**. The total European consumption of LAS alone was estimated to be about 0.43 million tons in the year 2005 **[187]**. LAS are commercially available as a mixture of homologs with alkyl chains ranging from C_{10} to C_{13}. In these technical mixtures, the C_{11} and C_{12} homologs are dominating, which translates to a weighted average carbon number between 11.7–11.8 **[186]**. As the benzene sulfonate group may be attached to any internal carbon atom of the alkyl chain, each homolog contains five to seven positional isomers **[188]**.

Alkyl ethoxysulfates (AES, also known as alkyl ethoxylated sulfates, alcohol ethoxysulfates or alcohol ethoxylated sulfates) are another important class of anionic

surfactants. They are commonly used in various consumer products, such as shampoos, hand dishwashing liquids, and laundry detergents, as well as in industrial cleaning processes, as industrial process aids in emulsion polymerization, and as additives in the plastics and paint production. The total volume of AES used in Europe is estimated to be 0.28 million tons per year **[189]**. Commercially available AES are complex surfactant mixtures containing anionic homologs with alkyl chain lengths ranging from 8 to 18 carbon atoms. Each homolog can exhibit varying degrees of ethoxylation ranging from 0 to 9 EO units **[190]**. However, the majority of AES blends manufactured, are alkyl chains in the range of C_{12} to C_{15} with 0 to 4 EO units **[191]**. A high production volume example of AES is sodium lauryl ether sulfate (SLES, sometimes also named sodium laureth sulfate). SLES is the sodium salt of the C_{12} homolog of AES with predominantly three EO units **[190]**. It should not be confused with sodium dodecyl sulfate (synonymously sodium lauryl sulfate, SLS), which belongs to the group of non-ethoxylated alkyl sulfates (AS). In general, AS can account for up to 50% of a technical AES mixture, but are also produced and used separately from AES **[191]**.

Other important surfactants are *secondary alkane sulfonates* (SAS) which are used in the production process for dishwashing and laundry products. Non-ionic surfactants with widespread applications are *alcohol ethoxylates* (AEO) and *nonylphenol ethoxylates* (NPEO). The latter are known to be degraded to *nonylphenol* (NP) and its short-chain ethoxylates, which are endocrine-disrupting compounds **[192]**. *Polyethylenoglycols* (PEGs) are widely used in the production of pharmaceuticals, cosmetic products, lubricants, antifreeze mixtures, hydraulic fracturing fluids, and surfactants **[193, 194]**. Furthermore, the production and application of surfactants imply the simultaneous accruement of large quantities of chemicals used for their synthesis and by-products; e. g. commercial LAS mixtures usually contain about 15% of byproducts **[195]**. Studies have shown that modern surfactants are extensively removed by a combination of biodegradation and sorption/settling processes during wastewater treatment **[196, 197]**. This leads to the release of a complex mixtures of numerous surfactants and their TPs into receiving rivers. Since the high removal rates are compensated by the exceptionally high production volume of surfactants and their

continuous introduction into the environment, surfactants can be considered as “pseudo-persistent” contaminants. Therefore, they may be able to cause the same exposure potential as persistent contaminants and may pose a risk to nearby reaches.

Most studies on the occurrence and behavior of modern surfactants during wastewater treatment were conducted from the late 1980s to the early 2000s. In these studies, average concentrations were mostly derived for a small number of WWTPs. In the past, most publications covered a rather limited number of analytes, as target analyses was used, which inherently focuses on few pre-selected substances for which reference standards are available. Thus, monitoring did not provide information about the multitude of substances occurring in environmental samples even not about structurally related substances like TPs or about compounds of the same chemical class. Suspect screening by high-resolution mass spectrometry (HRMS), using lists of environmentally relevant substances, can be an effective way to partly overcome this limitation and to gain a better insight on the occurrence of emerging substances in environmental samples. This study aims to provide current status on the presence of LAS and AES residues in WWTP effluents, especially against the backdrop of the changing surfactant production in Europe over the last decades. Target methods were supplemented by wide-scope suspect screening in digitally archived chromatograms for substances of the same class and TPs in order to provide a profound base for an ecotoxicological risk assessment for surfactants.

7.2 Material and methods

7.2.1 Chemicals

A LAS standard containing a mixture of homologs (CAS-Number.: 68584-22-5; 97%) with an advertised average chain length of 11.4 was purchased from Alfa Aesar (Karlsruhe, Germany). AES (CAS-Number.: 9004-82-4; 70.5%) and sodium dodecyl-d_{25} sulfate (SDS-d_{25}; ≥98%) were obtained from Santa Cruz Biotechnology, Inc. (Heidelberg, Germany). Metformin (97%) and PEG-04 (CAS-Number: 112-60-7; 99%) was procured from Sigma-Aldrich (Steinheim, Germany). Tetraethylene glycol

monododecyl ether (CAS Number 5274-68-0; 98%) was purchased from Merck KGaA (Chalkidona, Greece). Both laboratories involved in the study used the identical analytical standards for the quantification of LAS and AES.

Methanol (MeOH), Acetonitrile (ACN), and formic acid (all LC-MS grade) were purchased from Honeywell (Seelze, Germany). Ultra-pure water (LC-MS grade) was procured from VWR International (Bruchsal, Germany). Glacial acetic acid (100%) was obtained from Merck (Darmstadt, Germany). Ammonium acetate (≥98%) was provided by Sigma-Aldrich (Steinheim, Germany). Ammonium fluoride (≥98%) was purchased from Carl Roth (Karlsruhe, Germany). ACN and MeOH for HRMS (both LC–MS grade) were purchased from Merck (Darmstadt, Germany) and formic acid (99%) was obtained from Sigma-Aldrich (Buchs, Switzerland). Distilled water was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV, Bedford, MA, USA).

7.2.2 Sampling sites and sample preparation protocol

Seven-day composite samples of wastewater treatment plant (WWTP) effluents were obtained from 33 conventional WWTPs across Germany, which predominantly receive domestic wastewater (**Figure S7-1, (SI)**). The population equivalents (PE) of the sampled WWTPs ranged from 1,000 to 1,300,000 PE. Sampling took place from February to April 2018. Samples were taken by automatic samplers, stored in 10 L stainless steel containers, and immediately frozen after sampling. A seven-day composite sample was obtained by combining seven consecutive 24-hour composite samples. After arriving at the laboratory, each seven-day composite sample was thawed at room temperature. Subsequently, an aliquot was transferred to a 50 mL polypropylene tube and stored at −18 °C until analysis. Almost all WWTPs were each sampled under normal operational conditions and during dry weather periods (for 32 WWTP dry weather and one was sampled once during wet weather conditions). Additionally, four influents were sampled at the corresponding WWTPs, using the identical sampling approach as used for the effluents.

7.2.3 Instrumental analytical methods

Ultra-high performance liquid chromatography electrospray tandem mass spectrometry (UHPLC-MS/MS) was used for target analysis of LAS and AES, whereas UHPLC coupled to quadrupole time-of-flight (UHPLC-QTOF) was used for suspect screening of the TPs of LAS and other known surfactants. The UHPLC-QTOF method was also used for cross validation of the results obtained by HPLC-MS/MS target analysis. The analysis included the TPs of LAS sulfophenyl alkyl carboxylic acids (SPACs) and sulfophenyl alkyl dicarboxylic acids (SPADCs), the LAS-byproducts di-alkyl tetralin sulfonates (DATSs), and the TPs of DATS sulfo-tetralin alkyl carboxylic acids (STACs) and sulfo-tetralin alkyl di-carboxylic acids (STADCs). Further analysis comprised NPEO, nonylphenol ethoxylate sulfate (NPEO-SO_4), SAS, glycol ether sulfates (GES), PEGs, and AEO (**Table 8**).

Details regarding sample preparation, analytical instrumentation, MS settings and chromatographic conditions (LC columns, buffers, gradient programs, etc.) of all methods used for this study can be found in **Section S7.2 (SI)**.

7.2.4 Method validation of target analysis

The precision and accuracy of the analytical methods for LAS and AES used for target analysis were determined by extracting six aliquots of a tap and a WWTP effluent sample, respectively. All samples were spiked with a total concentration of 100 µg/L for LAS and 25 µg/L for AES (sum of C12 and C14 with EO0–9) prior to extraction. To account for any possible background concentration in the native water samples, two non-spiked tap water and effluent samples were analyzed for LAS and AES, respectively. For HRMS measurements five-point calibration curves were generated using linear regression analysis. The linearity was qualified by the linear correlation coefficient, r^2. Accuracy of the method was assessed with recovery experiments in wastewater effluent samples. Regarding sensitivity, method limits of detection and quantification (LODs and LOQs) were calculated from the recovery experiments at the lowest spiked concentration.

7.2.5 Methods for interpreting “non-detect” data

Left-censored observations, sometimes referred to as “non-detects” or “less than” values (e.g. <10 ng/L), are concentrations that are known only to be somewhere between zero and the LOQ. A commonly used method in environmental chemistry, to deal with values below the LOQ, is to substitute a fraction of the LOQ for each censored value, or to exclude them from the analysis. However, in recent years research has shown that this approach produces poor estimates of statistics, such as means, correlation coefficients, regression slopes, or hypothesis tests and can obscure trends or other patterns in the data **[198, 199]**. Better methods for interpreting censored values include regression on order statistics (ROS) and maximum likelihood estimation (MLE). ROS was used for a better estimate of average concentrations and for censored boxplots. The applied techniques are described in Helsel (2006) and are implemented in the NADA-package for R and were applied for the LAS and AES target analysis within this study.

7.2.6 Wide-scope suspect screening and data processing

HRMS chromatograms were recalibrated using HPC fitting algorithm, which is embedded in DataAnalysis 4.3. (Bruker Daltonics, Bremen, Germany). The manufacturer’s calibration method ensures mass accuracy below 2 mDa throughout the chromatographic run for *m/z* from 50 to 1200 Da. For exporting files in the mzML format, CompassXport 3.0.9.2. (Bruker Daltonics, Bremen, Germany) was used. Chromatograms acquired under data-independent acquisition were separated in low and high collision energy layer chromatograms.

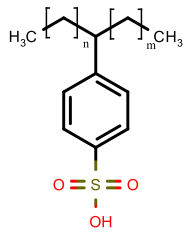
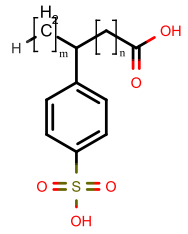
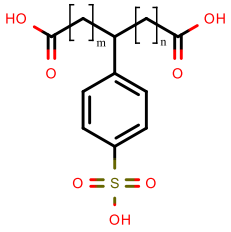
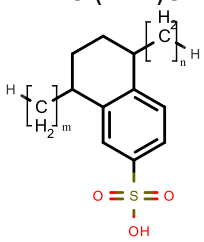
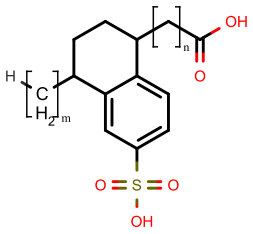
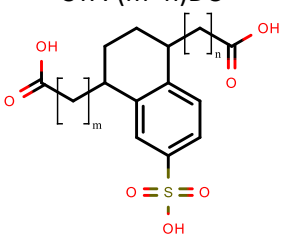
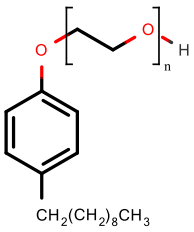
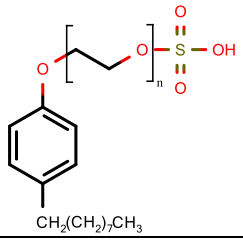
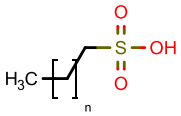
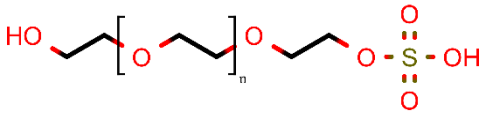
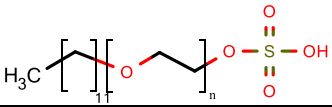
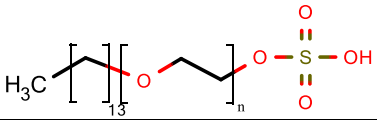
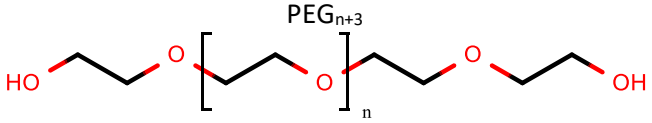
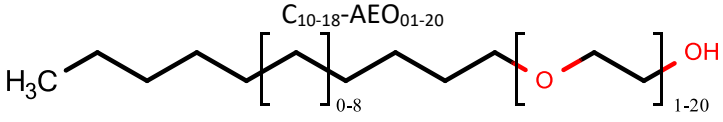
All mzML files and their metadata (instrumental, sample metadata, matrix-specific metadata and retention time of retention time index (RTI) mixture **[142]**), were uploaded to a separate section of the NORMAN Digital Sample Freezing Platform (DSFP) **[200]**, which has an in-built integrated standard operating procedure (SOP) to process the mzML files and all metadata for an automated generation of an Excel-

based Data Collection Templates (DCTs). DSFP was used to screen the results which were further evaluated and are visualized at a website (www.norman-data.eu/EWW_GERMANY) in order to show the spatial distribution of the analyzed surfactants.

In the first step, HRMS was used to confirm the concentrations of AES and LAS obtained by UHPLC-MS/MS. Suspect screening was then applied to search for the presence of TPs of LAS and other known surfactants expected to be present in wastewater effluents. All surfactants currently enlisted in NORMAN Suspect List Exchange (more specifically: the lists S7 EAWAGSURF, S8 ATHENSSUS and S23 EIUBASURF) were screened [137]. These lists have been compiled after a systematic literature review **[39, 52, 201-207]** by research groups within the NORMAN network **[13]**. A summary of chemical structures screened for in the samples is given in **Table 8**. The lists available in the NORMAN SusDat were extended for screening for PEGs with a higher number of ethoxy groups $(CH_2CH_2O)_x$, since it was found that PEGs with a higher mass are ionized in positive ionization as $[M+NH_4]^{2+}$.

SPACs, SPADCs, DATSs, STACs and STADCs were semi-quantified based on the comparison of their signals to the LAS surfactants. PEGs were semi-quantified based on PEG-04 for which an analytical standard was available. AEOs were semi-quantified based on the calibration curve of tetraethylene glycol monododecyl ether.

Table 8. General structures of linear alkylbenzene sulfonates (LAS), sodium alkyl ethoxysulfates (AES), LAS-related byproducts and TPs as well as other surfactants investigated within this study.

LAS	LAS_{m+n} 		
1st generation metabolites of LAS	SPA-(m+n)C 	SPA-(m+n)DC 	
2nd generation metabolites of LAS	DATS-(m+n)C 	STA-(m+n)C 	STA-(m+n)DC 
NPEO and NPEO-SO4	$NPEO_n$ 	$NPEO_n$-SO_4 	
SAS and GES	C_{n-2}-SAS 	GES_{n+1} 	
AES	C12-AES_n 	C14-AES_n 	
PEGs	PEG_{n+3} 		
C_{10-18}-AEO_{01-20}	C_{10-18}-AEO_{01-20} 		

7.3 Results and discussion

7.3.1 Method development of target analysis

Since the concentrations of individual LAS homologs in the standard were unknown, an experimental determination of individual homolog concentrations using single MS was performed. For this approach the following assumption has to be made: The response of the detector is identical for every homolog, which means that the intensity (I_n) of the mass spectrum measured by single MS for homolog n is directly proportional to its molar concentration (c_n) and B is a constant to all homologs (equation 1):

$$I_n = B \times c_n \quad \text{with} \quad c_n = \beta_n / M_n \tag{1}$$

As the total concentration of the LAS standard (β_{tot}) is known, the concentration of an individual homolog in the stock solution can be estimated by using the intensity measured by MS in counts per second, weighted by its molecular weight (MW) in Dalton. The calculations were based on data from multi-channel acquisition (MCA) scans. The results are displayed in **Table S7-2A**. As reported in previous studies **[208, 209]**, the C_{11}-LAS and C_{12}-LAS homologs also dominated the analytical standard, with 39% and 35%, respectively. C_{14}-LAS was not present in the standard. The calculated average chain length of the standard based on the experimental determination was 11.4, which is in accordance with the average number provided by the manufacturer. The results regarding the distribution of homologs fit well to those reported for LAS used in Europe and the USA **[186]**.

For AES a similar approach was used as for LAS. The experimental determination of individual concentrations of AES homologs/ethoxymers was based on the mass spectra obtained by direct injection of the AES standard in negative ionization mode. The most abundant ethoxymer for C_{12} and C_{14} was the one with zero EO units. A declining tendency in abundance was observed with an increasing number of EO units. Since other AES homologs were not visible in the mass spectra in negative ionization mode, it was assumed that the obtained standard only contains AES-C_{12} and AES-C_{14} homologs. The results are summarized in **Table S7-2B**. The AES standard contained

72% of C_{12}-AES and 28% of C_{14}-AES. The average number of EO units of C_{12}-AES in the standard was 2.5, which is in accordance with the reported average number of 2.7 for AES for domestic use and 2.4 for the total AES produced **[189]**.

A variety of reversed-phase LC columns and eluents were tested during method development. When using ACN as organic solvent, retention and peak shapes of LAS homologs were generally poor. In fact, LAS homologs had a similar retention as metformin, which was used to estimate the void time of the analytical method. MeOH was subsequently chosen as the organic mobile phase (eluent B), as it enabled sufficient chromatographic separation as well as good peak shapes. By applying a less retentive C_8 column, isomers of each LAS eluted as single peaks which facilitated peak integration. Ammonium formate (0.1 mM) in the aqueous mobile phase (eluent A) was found to be an effective mobile phase additive to increase the peak intensities of LAS homologs.

For AES the best performance in terms of chromatographic resolution, peak shape, intensity, and run time was achieved using the Luna Omega Polar C_{18} column with ultra-pure water (25 mM ammonium acetate, pH 3.6 adjusted with glacial acetic acid) (A) and acetonitrile (B) as eluents. The method enabled good peak shapes, high intensities, and short runtime.

During method development, considerable peaks of LAS in zero volume injections were observed. The contamination originated from the mobile phase(s) and it was found that a low organic content at the beginning of the gradient led to an accumulation of LAS on the analytical column during the post-run column equilibration, which was subsequently eluted during the next run. However, by starting with a MeOH content higher than 40%, the LAS peaks in zero volume injections were substantially decreased. Contamination issues in the laboratory with surfactants can be explained by their excessive usage in personal care products as well as in detergents and have already been reported by several authors **[210-213]**. Due to the issue of LAS contaminated solvents and laboratory equipment, a rotational vacuum concentrator was chosen for the extraction process. This technique was favored over solid phase extraction as the sample vessels can be heated to 550 °C to remove any potential surfactant residues (which is not possible when using the SPE

manifold and cartridges) and all samples including the calibration standards could be reconstituted in the same way.

Sample-pretreatment for AES was carried out the same way as for LAS using a rotational vacuum concentrator. However, the sample volume was increased to 10 mL, in order to account for the lower concentrations of AES compared to LAS in effluent samples. After evaporating the sample to dryness, the analytes were reconstituted in 1 mL of ultra-pure water:ACN (50:50, v:v).

For LAS, relative standard deviations (RSDs) ranged from 3% to 10% for both, tap and effluent water samples. For AES, RSDs ranged from 1% to 6% (tap water) and 3% to 10% (effluent water). Recoveries ranged from 91%–107% and 98%–115% for tap and effluent water samples for LAS, respectively. For AES, recoveries were between 94% and 114% and 90% and 120% for tap and effluent water samples, respectively. Both methods showed very good linearity within the calibration range. The LOQ of each homolog/ethoxymer was determined according to DIN 32645. For HRMS screening, calibration curves obtained were linear for wide concentration ranges with r^2>0.98 in all cases. Extraction recoveries for most target analytes showed recovery efficiency between 70% and 110%. To ensure a correct quantification, method precision was determined as %RSD from the recovery experiments and a precision limit <15% RSD was met for all target analytes.

7.3.2 LAS and AES in effluents and their removal efficiency

As the obtained data include left-censored values, ROS was used for a better estimate of average concentrations, and to draw censored boxplots. Individual LAS concentrations were in the lower to mid µg/L-range for the monitored WWTP effluents. Individual concentrations of C_{10}-LAS ranged from <LOQ to 18 µg/L with an estimate average concentration of 4 µg/L. With one exception, individual concentrations of C_{11}-LAS were consistently the highest among all LAS homologs at the same sampling point, with concentrations ranging from <LOQ to 20 µg/L and with an estimated average concentration of 5.4 µg/L. For C_{12}-LAS, individual concentrations ranged from <LOQ to 11 µg/L and an estimate average concentration of 3.4 µg/L. The lowest individual concentrations were measured for C_{13}-LAS, ranging between <LOQ

and 5.2 µg/L with an estimated average of 1.6 µg/L. Total LAS concentrations ranged from <LOQ to 47.7 µg/L. Based on the estimated average concentrations of individual LAS homologs, the average total LAS concentration in monitored WWTP effluents was 14.4 µg/L. The average LAS chain length of effluent samples was 11. **Figure 14** shows censored boxplots of the concentrations of LAS homologs in effluent samples. The highest median concentration was observed for C_{11}-LAS and the lowest for the C_{13} homolog, which also had the highest number of measurements below the LOQ. The boxplots further demonstrate that more than half of the observations for each homolog were within one order of magnitude, indicating that effluent concentrations of monitored WWTPs were similar to each other.

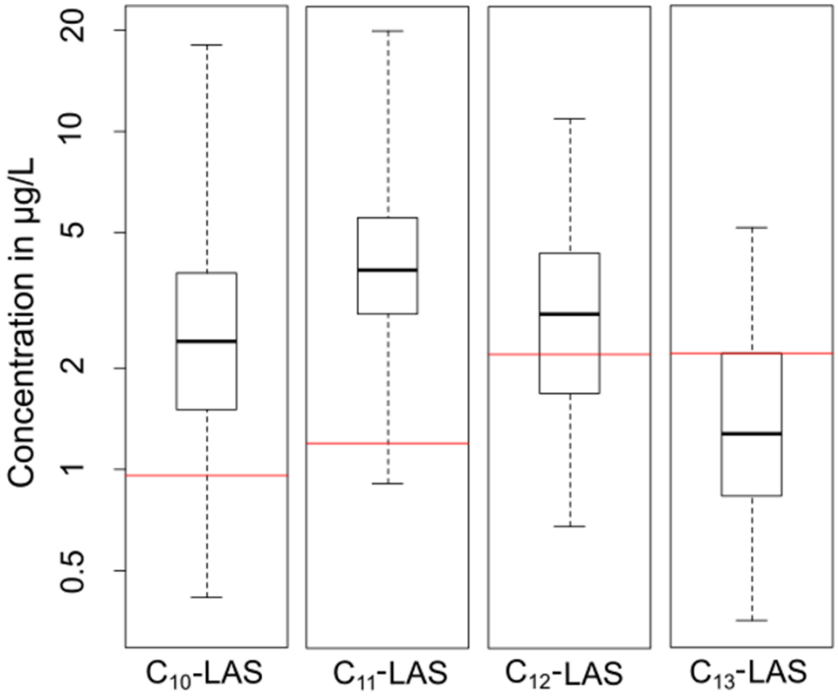


Figure 14. Censored boxplots of the concentrations of LAS homologs in effluents of monitored WWTPs. The horizontal red lines depict the limit of quantification for the respective homolog.

The average LAS chain length of 11.2 is in accordance with the average of 11.3 reported in McDonough et al. (2016) **[213]**. Since the average chain length in commercial products is between 11.7 and 11.8 **[186]**, this indicates the preferential removal of long alkyl chains during wastewater treatment. This could be explained by

the higher affinity of LAS homologs with long alkyl chains for suspended solids and sediments **[214]**. However, higher average chain lengths of LAS in the effluent of 11.6 **[215]** and 12.1 **[191]** have also been reported previously. The measured concentration of LAS homologs in WWTP effluents analyzed in this work were in some cases considerably lower than those reported in other studies. For example, the estimated average total LAS concentration of 14.4 µg/L for WWTP effluents in this study was about ten times lower than the findings for various WWTPs in Spain **[216, 217]**. However, when comparing effluent concentrations of different WWTPs, it is important to always consider the corresponding influent concentrations and removal rates. For example, it has been found that the removal of LAS in WWTPs equipped with trickling filters was more variable and overall lower, than in WWTPs using the activated sludge process **[191, 218]**. In the here presented study, total LAS concentrations of four influent samples ranged between 2,600 µg/L and 3,500 µg/L, which translates to very high removal rates between 99.2% and 99.8%. In contrast, average removal rates for LAS in the aforementioned studies by Riu et al. (2000) and Lunar et al. (2006) were only 83.6% and 87.7%, respectively, leading to elevated concentrations of LAS homologs in the effluent **[216, 217]**. Other authors reported average removal rates and effluent concentrations similar to the values determined in the present study. At nine WWTPs in Austria the average effluent concentration was 13.3 µg/L with an average removal rate of 99.7% **[219]**. For six WWTPs in the Netherlands the average effluent concentration was 43 µg/L with an average removal of 99.2% **[215]**. In a recently published study of effluent concentrations of 44 WWTPs in the U.S. the mean outflow concentration was 15.3 µg/L. However, no influent concentrations or removal rates were reported **[213]**.

The estimated average total AES concentration of 0.57 µg/L in WWTP effluents found in this work (**Figure 15**) was lower compared to values reported in other studies. McAvoy et al. (1998) determined average total AES effluent concentrations (28 analytes: C_{12}-C_{15} with $EO_{0–6}$) of 11 µg/L and 73 µg/L for activated sludge (average removal: 98%) and trickling filter treatment (average removal: 83%), respectively **[191]**. An average effluent concentration of 6.5 µg/L for AES (36 analytes: C_{12}–C_{15} with EO0–8) with a removal greater 99% was reported for effluents from seven WWTP in

the Netherlands **[215]**. In the study of McDonough et al. (2016) on 44 WWTPs in the U.S. the average total AES concentration (20 analytes: C_{12}–C_{16} with EO1–4) was 1.95 µg/L. One possible explanation for the overall low average total AES concentration in the present study, is that only homologs with an alkyl chain length of 12 and 14 were considered for the calculation of total AES concentrations. In the effluent of a trickling filter plant in the U.S. sampled by McAvoy et al. (1998) C_{12} and C_{14} homologs only accounted for 57% of the total AES concentration, while C_{13} and C_{15} homologs represented 43% **[191]**. No correlation between total LAS and AES effluent concentrations was found for the WWTPs monitored in the present study, indicating regional variations in the surfactant use and/or differences in the removal mechanisms of surfactants in the sewer and the WWTP. Concentrations of specific surfactants are further dependent on the respective inflow concentrations, which in turn are foremost controlled by regional differences in the per-capita surfactant use.

The comparison of results obtained by the two chemical laboratories involved in the study showed a good agreement considering the variability of the subsample, the variability introduced by the different sample preparation techniques applied and the different instrumental facilities. In all cases, concentration levels were at the same order of magnitude. A very good agreement was achieved for the LAS surfactants (e. g. deviation below 39% for C_{11}-LAS). Concentration levels of detected AES surfactants were in the low-ng/L range (close to the LODs of both methods), which could explain higher deviation between the results.

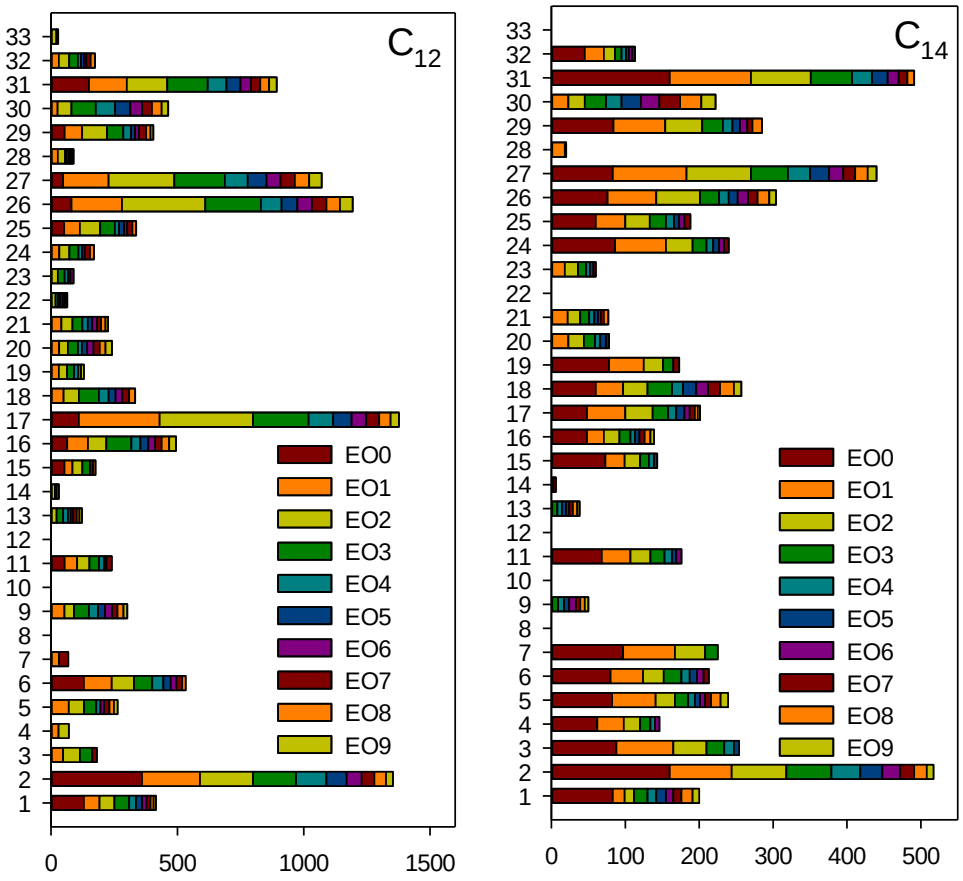


Figure 15: Concentrations for AES C_{12} (left) and AES C_{14} (right) with 0 to 9 ethoxy units in 33 WWTP effluents.

7.3.3 Suspect screening of other LAS-related by-products and transformation products

Interesting findings were revealed for TPs of LAS surfactants, DATSs and its TPs (**Table 9** for SPACs and SPADCs; **Table S7-3A** for DATSs; **Table S7-3B** for STACs and STADCs). Individual concentrations above the LOQ were summed up to calculate the total concentrations of DATSs, SPACs and STACs at each sampling point. They were determined at high concentration levels, whereas STADC and SPADC remained undetected. The highest total concentration was observed for DATSs (19 µg/L) followed by SPACs (17 µg/L) and STACs (5.3 µg/L). The sum of the concentrations of all by-products and TPs surpassed the concentration of LAS in most of the cases. In all cases, both the lower and higher mass homologs remained undetected, while medium

mass homologs were detected at high concentration levels. For example, SPA-1C, SPA-2C, SPA-3C and SPA-14C, SPA-15C remained undetected and maximum concentration levels were observed for medium mass homologs (SPA-8C, SPA-9C, SPA-10C and SPA-11C for SPAC; STA-5C and STA-6C for STAC and C10-DATS).

Table 9: Occurrence of metabolites of LAS (SPACs and SPADCs) in wastewater effluent samples. Semi-quantification was based on calibration curve of LAS-C_{10}. Concentrations are in ng/L. N.D.: not detected.

WWTP	SPA-1C	SPA-2C	SPA-3C	SPA-4C	SPA-5C	SPA-6C	SPA-7C	SPA-8C	SPA-9C	SPA-10C	SPA-11C	SPA-12C	SPA-13C	SPA-14C	SPA-15C	SPA-0-15DC	Total SPAC
1	N.D.	N.D.	N.D.	26	79	240	720	1,700	1,900	2,400	1,400	630	110	N.D.	N.D.	N.D.	9,205
2	N.D.	N.D.	N.D.	N.D.	18	78	190	410	690	770	510	200	55	N.D.	N.D.	N.D.	2,921
3	N.D.	N.D.	N.D.	N.D.	31	96	320	880	1400	2,000	1,400	690	130	N.D.	N.D.	N.D.	6,947
4	N.D.	N.D.	N.D.	11	46	140	470	930	970	1,100	730	390	95	N.D.	N.D.	N.D.	4,882
5	N.D.	N.D.	N.D.	21	58	170	620	1,200	1100	1,500	970	480	97	N.D.	N.D.	N.D.	6,216
6	N.D.	N.D.	N.D.	N.D.	60	240	670	1,100	1,500	2,000	1,300	730	200	N.D.	N.D.	N.D.	7,800
7	N.D.	N.D.	N.D.	20	87	310	1,300	3,300	3,600	4,100	2,000	830	130	N.D.	N.D.	N.D.	15,677
8	N.D.	N.D.	N.D.	N.D.	9.2	36	40	100	200	260	180	110	N.D.	N.D.	N.D.	N.D.	935
9	N.D.	N.D.	N.D.	N.D.	35	120	230	550	790	910	610	320	75	N.D.	N.D.	N.D.	3,640
10	N.D.	N.D.	N.D.	N.D.	26	110	250	750	1,100	1,200	800	380	85	N.D.	N.D.	N.D.	4,701
11	N.D.	N.D.	N.D.	12	39	130	560	900	860	930	570	270	58	N.D.	N.D.	N.D.	4,329
12	N.D.	N.D.	N.D.	N.D.	21	68	160	350	490	600	460	170	55	N.D.	N.D.	N.D.	2,374
13	N.D.	N.D.	N.D.	21	67	210	630	1,500	1,500	1,800	1,100	480	74	N.D.	N.D.	N.D.	7,382
14	N.D.	N.D.	N.D.	33	43	170	510	1,300	1,500	1,900	1,300	600	91	N.D.	N.D.	N.D.	7,447
15	N.D.	N.D.	N.D.	N.D.	64	190	840	2,700	2800	3,600	1,900	960	190	N.D.	N.D.	N.D.	13,244
16	N.D.	N.D.	N.D.	N.D.	N.D.	64	150	390	500	610	460	220	75	N.D.	N.D.	N.D.	2,469
17	N.D.	N.D.	N.D.	N.D.	52	280	740	1,100	1,100	1,100	750	330	86	N.D.	N.D.	N.D.	5,538
18	N.D.	N.D.	N.D.	N.D.	N.D.	46	160	330	510	710	550	250	61	N.D.	N.D.	N.D.	2,617
19	N.D.	N.D.	N.D.	16	57	270	660	1,300	1,400	1,600	1,300	430	70	N.D.	N.D.	N.D.	7,103
20	N.D.	N.D.	N.D.	N.D.	N.D.	32	700	180	130	94	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	1,136
21	N.D.	N.D.	N.D.	N.D.	63	300	1,300	4,300	4,200	4,100	2,000	860	19	N.D.	N.D.	N.D.	17,142
22	N.D.	N.D.	N.D.	N.D.	49	210	540	1,200	1,300	1,300	880	440	110	N.D.	N.D.	N.D.	6,029
23	N.D.	N.D.	N.D.	8.2	32	74	170	500	710	1,000	830	410	80	N.D.	N.D.	N.D.	3,814
24	N.D.	N.D.	N.D.	N.D.	32	120	340	770	980	1,100	750	340	78	N.D.	N.D.	N.D.	4,510
25	N.D.	N.D.	N.D.	6.6	22	76	240	600	810	1,000	730	370	110	N.D.	N.D.	N.D.	3,965
26	N.D.	N.D.	N.D.	N.D.	54	130	320	900	1100	1,400	990	430	110	N.D.	N.D.	N.D.	5,434
27	N.D.	N.D.	N.D.	N.D.	39	130	350	790	1000	1,200	690	300	99	N.D.	N.D.	N.D.	4,598
28	N.D.	N.D.	N.D.	N.D.	31	200	340	980	1300	1,400	920	450	130	N.D.	N.D.	N.D.	5,751
29	N.D.	N.D.	N.D.	N.D.	31	100	280	700	980	1,300	940	460	98	N.D.	N.D.	N.D.	4,889
30	N.D.	N.D.	N.D.	N.D.	18	61	160	410	640	720	410	140	N.D.	N.D.	N.D.	N.D.	2,559
31	N.D.	N.D.	N.D.	18	31	91	390	1,000	1,100	1,300	800	370	73	N.D.	N.D.	N.D.	5,173
32	N.D.	N.D.	N.D.	N.D.	22	69	210	440	620	730	480	260	67	N.D.	N.D.	N.D.	2,898
33	N.D.	N.D.	N.D.	N.D.	32	120	240	500	940	1400	930	460	110	N.D.	N.D.	N.D.	4,732

7.3.4 Suspect screening of PEGs and $C_{10\text{-}18}$-$AEO_{01\text{-}20}$ surfactants

It is known from the literature that they result in $[M+NH_4]^+$ adducts instead of $[M+H]^+$ under electrospray ionization **[23, 52]**. In this study, 41 PEG compounds with repeating ethoxy groups were screened in the positive ionization mode (**Figure 16**). It was discovered that PEGs with high molecular mass are ionized as $[M+NH_4]^{2+}$ adducts, which resulted in the positive detection of the longest homolog series so far reported in the literature in effluent wastewater samples.

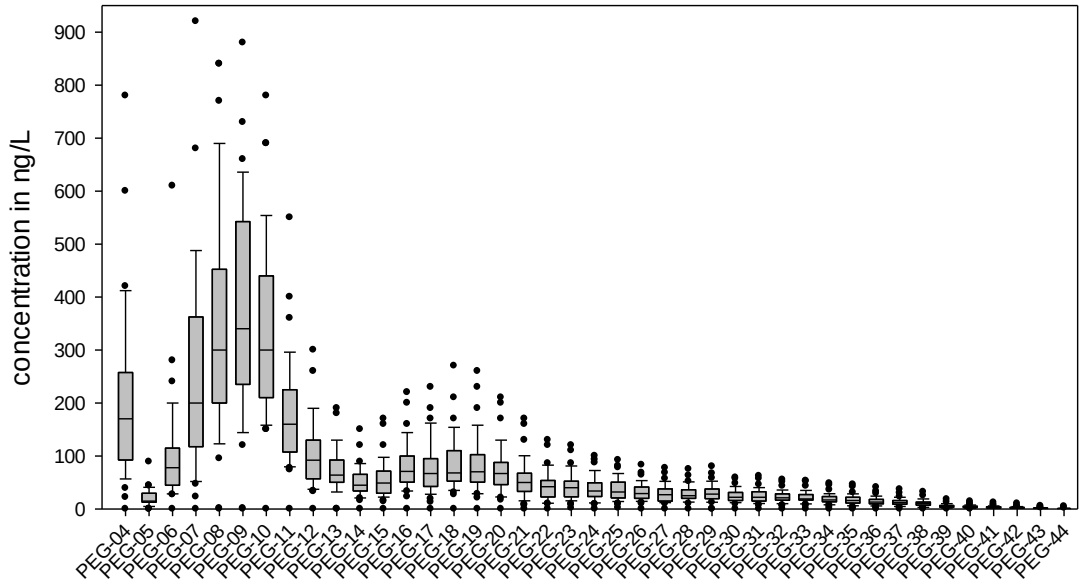


Figure 16. Distribution of PEGs across the 33 sampled WWTPs. Upper and lower limit of the box spans the first quartile to the third quartile, line in the box represents the median value, whiskers indicate the 5 and 95 percentiles with every outlier plotted as individual point.

The cumulative concentration level of all PEGs together was between 3.5 ng/L to 7,360 ng/L. This concentration level ranks PEGs as high as that of SPACs and STACs. Maximum concentration levels occurred in most of the cases for PEG-08 and PEG-09, followed by PEG-10 and PEG-07. PEGs were efficiently removed during biological wastewater treatment. However, they can also be generated during wastewater treatment when precursor molecules are biologically degraded. For example, low molecular PEGs have been described as the main metabolites of the nonionic

surfactants AEOs **[220, 221]** and NPEOs **[222]**. In an aerobic biodegradation test under OECD 301 test conditions, PEGs biodegraded more slowly than the parent AEOs and were removed by hydrolysis, thus leading to shorter PEG oligomers, and by oxidative hydrolysis, thus forming carboxylated PEGs **[223]**. Samples were screened for a total of 290 homologs of AEOs (C_8-C_{18}). This substance class also showed remarkable occurrence in the effluent wastewater samples (**Table S7.3C**). Homologs with medium ethoxy group content generally showed higher frequencies of appearance and concentrations.

7.3.5 Suspect screening of other surfactants with known fragmentation

A number of other surfactants that are shown in **Figure 17** have previously been reported in the literature and could not be semi-quantified here due to lack of standards with similar structure. However, their fragmentation pattern was known and thus they were identified at the level of 'possible structure by library spectrum match' (Level 2A; **[27]**). High frequency of appearance (FoA) was observed for C_{12}-SAS and C_{14}-SAS, which were detected in all wastewater samples, C_{11}-SAS was detected with FoA 91%, C_{13}-SAS was detected with FoA 73%. C_{10}-SAS and C_{16}-SAS were detected in only two wastewater effluent samples, while the rest of SAS surfactants remained undetected. The highest signal was observed for C_{12}-SAS.

Other surfactants with widespread occurrence were NP1ethoxycarboxylate, naphthalene-1-sulfonate and dihexyl sulfosuccinate (DHSS), which were detected with very high FoA (100% for NP1ethoxycarboxylate and naphthalene-1-sulfonate, 97% for DHSS). On the contrary, the WFD priority substance 4-nonylphenol (EQS in inland surface water 300 ng/L) was scarcely detected (FoA 12%) and at low concentration levels (<10 ng/L to 54 ng/L). However, its homolog compound NPEO1 was detected in almost all wastewater samples (FoA 97%). NPEO3 was detected only in a few samples (FoA 18%), while NPEO4 was found in more than half of the samples (FoA 58%). The rest of NPEO compounds (NPEO5–NPEO17) remained undetected.

4-Octylphenol, another WFD priority substance with EQS for inland surface waters of 100 ng/L, was frequently (FoA 85%) detected in the wastewater effluent samples in the concentration range from <17 ng/L to 170 ng/L. Its homolog substances OP1

ethoxy carboxylate and OP2 ethoxy carboxylate were also detected in 73% and 18% of the samples, respectively. GES surfactants were detected less frequently. Maximum FoA was observed for GES9 (FoA 55%), followed by GES11 (FoA 39 %) and GES10 (FoA 36 %).

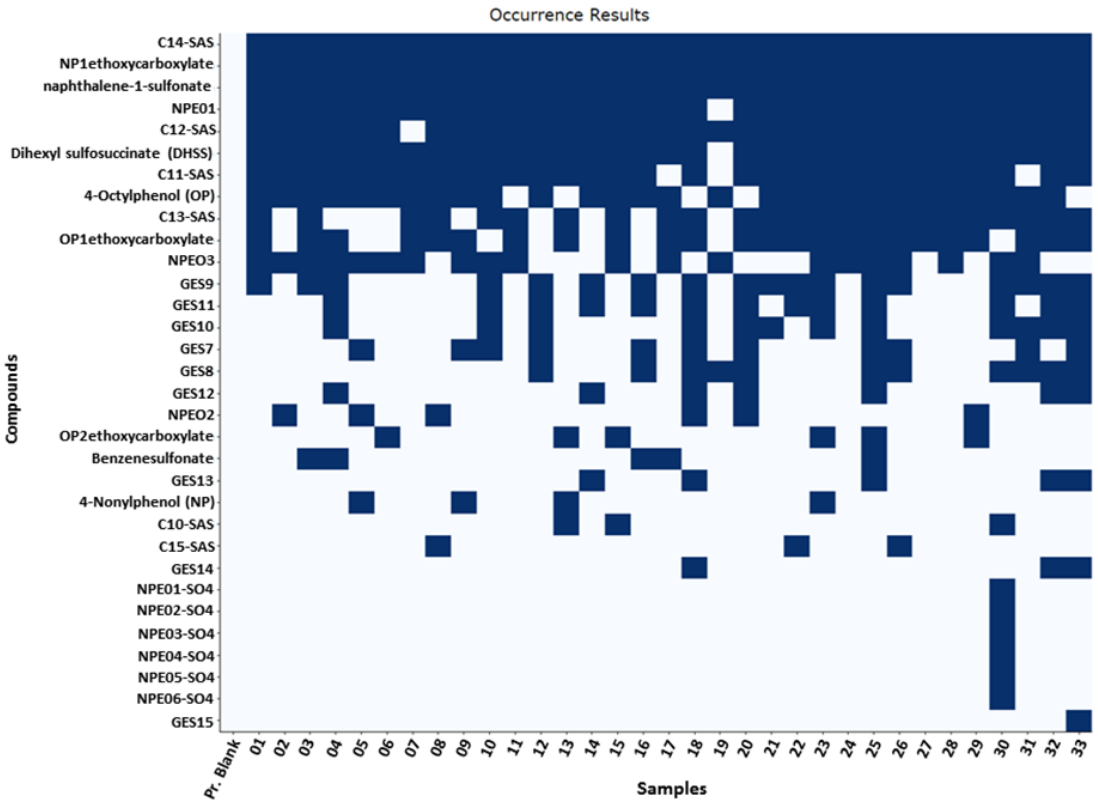


Figure 17. Occurrence profile of surfactants included in the two suspect surfactant lists (EAWAGSURF, ATHENSSUS, EIUBASURF), which were screened for their presence in the wastewater effluents samples using DSFP. Dark blue color indicates positive detection of a substance in the particular sample.

7.3.6 Suspect screening of surfactants using in silico predicted fragmentation

After the compilation of the EIUBASURF list, *in silico* predicted fragmentation patterns **[30]** of the candidate suspect compounds were generated. The list and the samples were uploaded to DSFP and wastewater effluent samples were screened for these substances. Substances with match of more than three *in silico* predicted fragments were prioritized and then further investigated. The investigation involved the

acquisition of HRMS/MS spectra and structural explanation of the spectra (procedure termed as 'annotation'). Candidates that could adequately explain the HRMS/MS spectra were summarized in **Figure** Error! Reference source not found.**S7-3D**. The c olor coding in indicates the number of fragments explained. Structures of tentatively identified compounds can be found in **Table S7-3E**. All these compounds were investigated in-depth, following the NTS identification workflow **[23]**, and were tentatively identified (Level 3, **[27]**). The presence of 1*H*-benzotriazole, propafenone and benzoic acid in the samples could be successfully confirmed with authentic standards (Level 1). Benzoic acid and benzotriazole were detected in all samples, while propafenone was detected with FoA 48%. Compounds with widespread occurrence were mono-C_{12} alkyl sulfosuccinate and lauroyl sarcosinate (FoA 100%), di-2-ethylhexyl sulfosuccinate and C_8-alkyl sulfate (linear) (FoA 91%), "amines, tallow, 1+2 EO (R=CH3)" and "amines, tallow, 5+5 EO (R=H)" (88%), cumene sulfonate (FoA 85%), C_{12} alkyl phosphate esters (FoA 7%), panthenol (FoA 67%) and methylparaben (55%). Compounds detected in less than half of the samples were C_{10} alcohol, predominately linear, 2 EO (FoA 48%), succinic acid (36%) and glycerides, C_{15} mono (27%). Finally, sulfates were detected only scarcely (C_{16}-alkyl 4 ethyl sulfate with FoA 18%, C_9-alkyl 2 ethyl sulfate with FoA 15% and C_8-alkyl 2 ethyl sulfate with FoA 6%).

Two examples of such in-depth investigations to tentatively identify compounds are shown in **Figure 18** (C_{12} alkyl phosphate ester) and **Figure S7-3F** (C_{14} alkyl dimethyl amine oxide). C_{12} alkyl phosphate ester was detected in both ionization modes and gave fragments of diagnostic evidence (e. g. 78.9591 and 96.9690 for negative and 98.9842 for positive ionization, respectively). C_{14} alkyl dimethyl amine oxide structure did not result in fragments of diagnostic evidence because of the structure of the compound. However, the obtained spectrum was clearly explainable, and all fragments could be annotated with respective structural fragments.

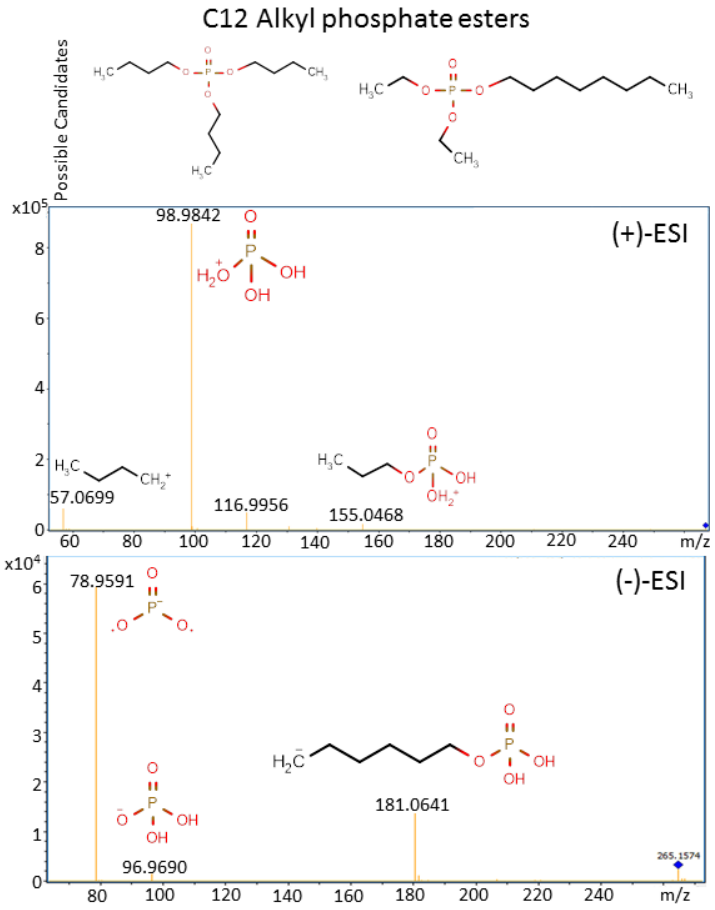


Figure 18. Tentative identification of C_{12} alkyl phosphate esters (level 3; ramification possible). Annotated fragment structures for positive and negative electrospray ionization (ESI) indicate that compound contains PO_4 moiety and a carbon chain.

7.4 Conclusions

Very high removal rates for LAS (>99.2%) and for AES (>99.8%) were observed, which confirms that both surfactants are extensively removed during conventional wastewater treatment. No correlation between total LAS and AES effluent concentrations was found for the WWTPs monitored in the present study, indicating regional variations in the surfactant use and/or differences in the removal mechanisms of surfactants in the sewer and the WWTP.

A screening of 1,564 surfactants and their metabolites in the effluent samples by UHPLC-ESI-QTOF-MS analysis showed that in many cases the sum of concentration of all LAS-related byproducts and TPs surpassed the concentration of LAS themselves; all

surfactants and surfactant-related compounds together accounted for concentrations up to 82 µg/L in a single sample. These are important findings for an updated ecotoxicological evaluation of surfactants.

An interactive map for visualization of concentrations of detected surfactants in the studied WWTPs is available at www.norman-data.eu/EWW_GERMANY. Based on the "Detergents Ingredients Database" (European Commission, 2016), a Suspect List of all collected surfactants and their metabolites was created including their exact masses and chemical structures (SMILES, InChIKey). It is available at www.norman-network.com/?q=node/236 as a list S23 coded 'EIUBASURF'. All UHPLC-ESI-QTOF-MS raw data were uploaded into NORMAN Digital Sample Freezing Platform (www.norman-data.eu) and they are available to for future retrospective screening of many more substances.

CHAPTER 8

CONCLUSIONS

The last decade, the introduction of HRMS instruments and the development of computational tools for data treatment of the HRMS data has revolutionized the field of environmental science. The development of generic sample preparation methods and instrumental analysis protocols provided unexplored opportunities to the researchers to tackle environmental problems and design holistic chemical monitoring programmes in a novel way. The presented thesis is a contribution towards these developments, which are necessary to accomplish successful environmental monitoring.

An open-source non-target screening workflow capable of detecting compounds with high fluctuation over time was applied in LC-HRMS data (*Chapter 3*). The critical parameters affecting the computational workflow were optimized and the open-source workflow was applied for analysis of influent wastewater to detect chemicals directly disposed to the sewage system and other relevant compounds. The prioritization workflow proved efficient and it was possible to elucidate top prioritized compounds. 18 compounds were tentatively identified and the occurrence of two new substances were revealed for the first time.

Also, in the context of this thesis and in cooperation with reference laboratories of the NORMAN network, a global decentralised early-warning system for revealing the presence of emerging substances in the environment was established. A joint pilot-scale activity was used to assess the feasibility of such a system. Eight reference laboratories participated and contributed 48 sets of samples from 14 countries and three continents. Laboratories were requested to investigate their data for the occurrence of 156 compounds. Suspect list included both established chemicals and newly-identified substances for which no literature occurrence data were available. A QA/QC scheme was developed and applied to evaluate and harmonise raw data and results from the participants. Within the pilot study it was possible to detect, discuss and propose solutions for critical points of the workflows that are frequently encountered by analytical chemists during the application of retrospective suspect screening.

The decrentralised early-warning system inspired the development of a web interface for sharing, archiving and screening of LC-HRMS data. The tool is termed Digital Sample Freezing Platform and enabled automated digital archiving and rapid and unbiased wide-scope suspect screening. DSFP incorporates major present advancements in HRMS, enables enhanced data mining and provides results visualization tools. DSFP was used to investigate the occurrence of 1447 substances in samples collected in the Black Sea (54 seawater, 19 sediment and 12 biota samples). The samples were screened for chemicals registered under REACH regulation and for antibiotics. Altogether, 80 compounds were detected in the samples. DSFP was further tested with additional data from the Black Sea and with data from two other large-scale monitoring programmes as described in *Chapters 6* and *7.*

Chapter 6 presents the results of effluent wastewater sampling campaign from samples collected at twelve WWTPs in nine countries (Romania, Serbia, Hungary, Slovenia, Croatia, Slovakia, Czech Republic, Austria, Germany). WWTPs' selection was based on countries' dominant technology and a number of served population with the aim to get a representative view of the pollution status. Wide-scope target screening resulted in detection and quantification of 280 compounds. Non-traditional monitoring tools such as *in vitro* bioassays and ARG analysis were applied parallel to chemical screening. Chemical risk assessment resulted in prioritization of 17 frequently occurring CECs exceeding toxicity threshold values(PFOS, Ofloxacin, Telmisartan, Diclofenac, C12-LAS, Carbamazepine, Ibuprofen, 4-tert-Octylphenol, Meclofenamic acid, Fipronil, Carbendazim, Venlafaxine, Clarithromycin, 4-Hydroxy-Omeprazole, EDDP, Temazepam and Sertraline). Additionally, a system to translate the signals from bioassays exceeding EBTs to an action plan at the level of WWTP's operatorswas proposed. A database was designed to enable storage and quick visualisation of the results.

Chapter 7 describes the monitoring of surfactants in effluent wastewater in Germany. In total 33 wastewater samples were collected with the purpose to be investigated for the occurrence of LAS and AES surfactants by target screening and get an overview of the occurrence of surfactants and their TPs. A list of all known surfactant substances was compiled by literature review and from the EU Detergents Ingredients Database. The suspect list was made available to researchers through the NORMAN Suspect list

exchange activity **[137]** and used to screen 1564 surfactants in the wastewater effluent samples using DSFP. 295 compounds were tentatively identified. Elevated concentration levels were observed for TPs of LAS (DATs, STAC, SPAC), the cumulative concentration of which surpassed in many cases the concentration of the parent LAS compounds. Remarkable was also the occurrence of AEO and PEGs. Concentration of all surfactants together in wastewater samples accounted for up to 82 μg/L.

In the context of this thesis, databases, visualization and chemometric tools were developed and applied for wide-scope screening and monitoring of emerging substances in various European ecosystems. It was demonstrated that these tools can harvest the wealth of information contained in LC-HRMS data. However, continuous improvement of these tools such, such as DSFP, and addition of modules that provide enhanced data processing capabilities remain a challenging topic. Initiating sampling campaigns to obtain a critical mass of raw mass chromatograms covering samples from all environmental compartments across Europe and beyond is of utmost importance. Integration of GC-APCI-HRMS and GC-EI-HRMS data in DSFP would be a significant upgrade towards a unified global platform for storing, viewing and screening of much wider analytical window of environmental pollutants.

ABBREVIATIONS AND ACRONYMS

1dbCES1	One double bond Chain Ethoxy surfactants
A	Antibiotics
ABS	Branched Alkylbenzene Sulfonates
ACN	Acetonitrile
AEO	Alcohol Ethoxylates
AEO-PEGs	Alcohol Ethoxylated-Polyethyleno Glycols
AES	Alkyl Ethoxysulfates
ANDI-MS or netCDF	Analytical Data Interchange Format for Mass Spectrometry
ANOVA	Analysis Of Variance
anti-AR CALUX®	Anti-androgenic activity
anti-PR CALUX®	Anti-progestin activity+B61
APCI	Atmospheric Pressure Chemical Ionization
AR	Antibiotic Resistance
ARGs	Antibiotic Resistant Genes
AS	Alkyl Sulfates
BDS	BioDetection Systems
CA	Cellulose Acetate
CALUX® bioassays	Chemical Activated Luciferase Expression bioassays
CAS	Chemical Abstracts Service
CECs	Contaminants of Emerging Concern
CES	Chain Ethoxy Surfactant
CESIO	European Committee of Organic Surfactants and their Intermediates
CID	Collision-Induced Dissociation
CMF	Competitive Fragmentation Modeling
DATs	Di-Alkyl Tetralin Sulfonates
DCT	Data Collection Template (excel spreadsheet)
DDA	Data-Dependent Acquisition
DHSS	Dihexyl Sulfosuccinate
DIA	Data Independent Acquisition
DID	Detergents Ingredients Database
DNA	Deoxyribonucleic Acid
DRB	Danube River Basin
DSFP	Digital Sample Freezing Platform
EBTs	Effect-based Trigger values
ECHA	European Chemicals Agency
EDA	Effect-Directed Analysis
EDDP	2-Ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine
EHS	Ethoxy Hydrogen Surfactants
EI	Electron Impact ionisation
EI	Electron Impact

EIC	Extracted Ion Chromatogram
EMBLAS project	Environmental Monitoring in the Black Sea
EO	Ethoxy group
EoE	Extent of PNEC Exceedance
EQS	Environmental Quality Standards
ERα CALUX®	Estrogenic activity
ESI	Electrospray ionization
EU	European Union
FoA	Frequency of Appearance
FoE	Frequency of PNEC Exceedance
FWHM	Full Width at Half Maximum
GC	Gas Chromatography
GC-MS	Gas Chromatography coupled to Mass Spectrometry
GES	Glycol Ether Sulfates
GES	Glycol Ether Sulfates
GR CALUX®	Glucocorticoid activity
HRMS	High Resolution Mass Spectrometry
IC	internal control
ICPDR	International Commission for the Protection of the Danube River
InChI	International Chemical Identifier
InChIKey	Compact Hashed Code derived from InChI
IT-FT	Linear ion trap-Orbitrap
ITNANTIBIOTIC	List of 670 antibiotics collected in context of innovative training network (ITN) Marie Curie Program ANSWER
JBSS	Joint Black Sea Survey
JDS2	Joint Danube Survey 2
JDS3	Joint Danube Survey 3
KWR	Water Cycle Research Institute KWR
LAS	Linear Alkylbenzene Sulphonate
LAS	Linear Alkylbenzene Sulfonates
LC	Liquid Chromatography
LC-HRMS	Liquid Chromatography coupled to High-Resolution Mass Spectrometry
LC-MS	Liquid Chromatography coupled to Mass Spectrometry
LOD	Limit of Detection
log K_{OW}	Logarithm of n-Octanol:Water Partition Coefficient
logP	Logarithm of Partition Coefficient
LOQ	Limit of Quantification
MCA	Multi-Channel Acquisition
MEBA	Multivariate Empirical Bayes Approach
MEC_{95}	95^{th} percentile measured experimental concentration
MeOH	Methanol

MLE	Maximum Likelihood Estimation
MRM	Multiple Reaction Monitoring
NKUA	National and Kapodistrian University of Athens
NORMAN network	Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances
NORMAN SusDat	NORMAN Suspectlist Database
NormaNEWS	NORMAN Early Warning System
NPEO	Nonylphenol Ethoxylates
NPEO	Nonylphenol
NPEO-SO_4	Nonylphenol Ethoxylate Sulfate
Nrf2 CALUX®	Oxidative stress bioassay
NSAIDs	Nonsteroidal anti-inflammatory drugs
NSD	Normalized Standard Deviation
NTS	Non-target screening
NTS CWG	Non-target screening cross-working group
OBI-WARP	Ordered Bijective Interpolated Warping
OECD	Organisation for Economic Co-operation and Development
PAH CALUX®	Polycyclic Aromatic Hydrocarbons CALUX®
PAHs	Polycyclic Aromatic Hydrocarbons
PCBs	Polychlorinated Biphenyls
PE	Population Equivalents
PEG	Polyethylene Glycol
PEG	Polyethylenoglycol
PFAS	Polyfluoroalkyl Surfactants
PFOS	Perfluorooctane Sulfonate
PFOS	Perfluorooctanesulfonic acid
PNEC	Provisional No Effect Concentration
PPARg CALUX®	Peroxisome proliferators gamma
PPARα CALUX®	Peroxisome proliferators alpha
P-PNEC	predicted PNEC
PXR CALUX®	Xenobiotic metabolism bioassay
QC	Quality Control
Q-FT	Quadrupole-Orbitrap
QI	Qualifier (Fragment) Ions
qPCR	Quantitative Polymerase Chain Reaction
QqQ	Triple quadrupole mass analyser
QSRR	Quantitative Structure (Chromatographic) Retention Relationship
QSTR	Quantitative Structure Toxicity Relationship
QTOF	Quadrupole-Time-Of-Flight mass analyser

Q-TOF	Quadruple time-of-flight
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
REs	Responsive Elements
ROS	Regression on Order Statistics
rRNA	Ribosomal Ribonucleic Acid
RSD	Relative Standard Deviation
RTI	Retention Time Index
SAS	Secondary Alkane Sulfonates
SI	Supporting Information
SLES	Sodium Lauryl Ether Sulfate
SLS	Sodium Lauryl Sulfate
SMILES	Simplified Molecular-Input Line-Entry System
SOP	Standard Operating Procedure
SPACS	SulfoPhenyl Alkyl Carboxylic acids
SPACs	Sulfophenyl Alkyl Carboxylic Acids
SPADCs	Sulfophenyl Alkyl Dicarboxylic Acids
SPE	Solid Phase Extraction
STACs	Sulfo-Tetralin Alkyl Carboxylic Acids
STADCs	Sulfo-Tetralin Alkyl Di-Carboxylic Acids
TBEP	Tris(2-butoxyethyl) phosphate
TOF	Time-Of-Flight mass analyser
TPs	Transformation Products
UATHTARGETS	University of Athens target list
UCY	University of Cyprus
UHPLC	Ultra High Pressure Liquid Chromatography
UHPLC-MS/MS	Ultra-high performance liquid chromatography electrospray tandem mass spectrometry
UHPLC-QTOF	Ultra-high performance liquid chromatography electrospray tandem mass spectrometry
UNDP	United Nations Development Programme
USA	United States of America
UVCBs	Unknown or Variable Composition, Complex Reaction Products and Biological Materials
WFD	Water Framework Directive
WHO	World Health Organisation
WWTPs	WasteWater Treatment Plants
XML	Extensible Markup Language
ΕΚΠΑ	Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

REFERENCES

[1] S.D. Richardson, Environmental mass spectrometry: emerging contaminants and current issues, Anal Chem, 84 (2012) 747-778.

[2] M. Gavrilescu, K. Demnerova, J. Aamand, S. Agathos, F. Fava, Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation, N Biotechnol, 32 (2015) 147-156.

[3] M. Cesen, D. Heath, M. Krivec, J. Kosmrlj, T. Kosjek, E. Heath, Seasonal and spatial variations in the occurrence, mass loadings and removal of compounds of emerging concern in the Slovene aqueous environment and environmental risk assessment, Environ Pollut, 242 (2018) 143-154.

[4] M. Petrovic, D. Barcelo, S. Perez, Comprehensive Analytical Chemistry: Analysis, Removal, Effects and Risk of Pharmaceuticals in the Water Cycle Occurrence and Transformation in the Environment, Elsevier, 2013.

[5] European Chemical Agency, List of registered substances under REACH regulation https://echa.europa.eu/information-on-chemicals/registered-substances (Last accessed 15 February 2019).

[6] B.I. Escher, K. Fenner, Recent advances in environmental risk assessment of transformation products, Environ Sci Technol, 45 (2011) 3835-3847.

[7] Z. Tousova, J. Froment, P. Oswald, J. Slobodnik, K. Hilscherova, K.V. Thomas, K.E. Tollefsen, M. Reid, K. Langford, L. Blaha, Identification of algal growth inhibitors in treated waste water using effect-directed analysis based on non-target screening techniques, J Hazard Mater, 358 (2018) 494-502.

[8] C.M. Manaia, J. Rocha, N. Scaccia, R. Marano, E. Radu, F. Biancullo, F. Cerqueira, G. Fortunato, I.C. Iakovides, I. Zammit, I. Kampouris, I. Vaz-Moreira, O.C. Nunes, Antibiotic resistance in wastewater treatment plants: Tackling the black box, Environ Int, 115 (2018) 312-324.

[9] European Commission, DIRECTIVE 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy, Official Journal of the European Union, 226 (2013) 1-17.

[10] European Commission, Commission implementing decision (EU) 2018/840 of 5 June 2018 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council and repealing Commission Implementing Decision (EU) 2015/495 Official Journal of the European Union, 141 (2018) 9-12.

[11] European Commission, Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 on environmental quality standards in the field of water policy, amending and subsequently repealing Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council, Official Journal of the European Union, 348 (2008) 84-97.

[12] R. Altenburger, S. Ait-Aissa, P. Antczak, T. Backhaus, D. Barcelo, T.B. Seiler, F. Brion, W. Busch, K. Chipman, M.L. de Alda, G. de Aragao Umbuzeiro,

B.I. Escher, F. Falciani, M. Faust, A. Focks, K. Hilscherova, J. Hollender, H. Hollert, F. Jager, A. Jahnke, A. Kortenkamp, M. Krauss, G.F. Lemkine, J. Munthe, S. Neumann, E.L. Schymanski, M. Scrimshaw, H. Segner, J. Slobodnik, F. Smedes, S. Kughathas, I. Teodorovic, A.J. Tindall, K.E. Tollefsen, K.H. Walz, T.D. Williams, P.J. Van den Brink, J. van Gils, B. Vrana, X. Zhang, W. Brack, Future water quality monitoring--adapting tools to deal with mixtures of pollutants in water resource management, Sci Total Environ, 512-513 (2015) 540-551.

[13] V. Dulio, J. Slobodnik, NORMAN-network of reference laboratories, research centres and related organisations for monitoring of emerging substances, Environ Sci Pollut Res Int, 16 Suppl 1 (2009) S132-135.

[14] NORMAN Network, Network of reference laboratories, research centres and related organisations for monitoring of emerging substances www.norman-network.net (Last accessed 15 February 2019).

[15] N.S. Thomaidis, A.G. Asimakopoulos, A.A. Bletsou, Emerging contaminants: A tutorial mini-review, Global NEST J, 14 (2012) 72-79.

[16] European Commision, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC OJ L 396 of 30 December 2006, Official Journal of the European Union, 136 (2007) 3-280.

[17] P. Rostkowski, P. Haglund, R. Aalizadeh, A.N. Alygizakis, N.S. Thomaidis, J. Beltran Arandes, P. Booij, H. Budzinski, A. Covaci, S. Grosse, R. Hindle, I.M. Ipolyi, K. Jobst, S. Kaserzon, The Strength in Numbers: Comprehensive Characterization of House Dust using Complementary Mass Spectrometric Techniques, Anal Bioanal Chem, Accepted.

[18] J. Wu, L. Zhang, Z. Yang, A Review on the Analysis of Emerging Contaminants in Aquatic Environment, Crit Rev Anal Chem, 40 (2010) 234-245.

[19] F.J. Santos, M.T. Galceran, Modern developments in gas chromatography–mass spectrometry-based environmental analysis, J Chrom A, 1000 (2003) 125-151.

[20] Scientific Instrument Services (SIS), Mass Spectral Libraries (NIST 17 and Wiley Libraries) https://www.sisweb.com/software/ms/wiley.htm (Last accessed 15 February 2019).

[21] Y. Pico, D. Barcelo, Transformation products of emerging contaminants in the environment and high-resolution mass spectrometry: a new horizon, Analytical and bioanalytical chemistry, Anal Bioanal Chem, 407 (2015) 6257-6273.

[22] European Commision, Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (Text with EEA relevance)

(notified under document number C(2002) 3044), Official Journal of the European Union, 221 (2002) 8-36.

[23] P. Gago-Ferrero, E.L. Schymanski, A.A. Bletsou, R. Aalizadeh, J. Hollender, N.S. Thomaidis, Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS, Environ Sci Technol, 49 (2015) 12333-12341.

[24] J. Hollender, E.L. Schymanski, H.P. Singer, P.L. Ferguson, Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go?, Environ Sci Technol, 51 (2017) 11505-11512.

[25] F. Hernandez, J.V. Sancho, M. Ibanez, E. Abad, T. Portoles, L. Mattioli, Current use of high-resolution mass spectrometry in the environmental sciences, Anal Bioanal Chem, 403 (2012) 1251-1264.

[26] A.A. Bletsou, J. Jeon, J. Hollender, E. Archontaki, N.S. Thomaidis, Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment, TrAC Trends Anal Chem, 66 (2015) 32-44.

[27] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ Sci Technol, 48 (2014) 2097-2098.

[28] 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.

[29] M.E. Dasenaki, N.S. Thomaidis, Multi-residue determination of 115 veterinary drugs and pharmaceutical residues in milk powder, butter, fish tissue and eggs using liquid chromatography-tandem mass spectrometry, Anal Chim Acta, 880 (2015) 103-121.

[30] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification, Metabolomics, 11 (2014) 98-110.

[31] HighChem Ltd, HighChem Ltd, Mass Frontier™ a software package for the management, evaluation, and interpretation of mass spectra www.highchem.com (Last accessed 15 February 2019).

[32] C. Ruttkies, E.L. Schymanski, S. Wolf, J. Hollender, S. Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation, J Cheminform, 8 (2016) 3.

[33] R. Aalizadeh, N.S. Thomaidis, A.A. Bletsou, P. Gago-Ferrero, Quantitative Structure–Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples, J Chem Inf Model, 56 (2016) 1384-11398.

[34] R. Aalizadeh, P.C. von der Ohe, N.S. Thomaidis, Prediction of acute toxicity of emerging contaminants on the water flea Daphnia magna by Ant Colony Optimization-Support Vector Machine QSTR models, Environ Sci Process Impacts, 19 (2017) 438-448.

[35] Y. Moriya, D. Shigemizu, M. Hattori, T. Tokimatsu, M. Kotera, S. Goto, M. Kanehisa, PathPred: an enzyme-catalyzed metabolic pathway prediction server, Nucleic Acids Res, 38 (2010) W138-143.

[36] J. Jaworska, S. Dimitrov, N. Nikolova, O. Mekenyan, Probabilistic assessment of biodegradability based on metabolic pathways: catabol system, SAR QSAR Environ Res, 13 (2002) 307-323.

[37] J. Wicker, T. Lorsbach, M. Gutlein, E. Schmid, D. Latino, S. Kramer, K. Fenner, enviPath--The environmental contaminant biotransformation pathway resource, Nucleic Acids Res, 44 (2016) D502-508.

[38] N.A. Alygizakis, S. Samanipour, J. Hollender, M. Ibáñez, S. Kaserzon, V. Kokkali, J. Van Leerdam, J.F. Mueller, M. Pijnappels, R. M.J., E.L. Schymanski, J. Slobodnik, N.S. Thomaidis, K.V. Thomas, Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry, Environ Sci Technol, 52 (2018) 5135-5144.

[39] N.A. Alygizakis, P. Gago-Ferrero, J. Hollender, N.S. Thomaidis, Untargeted time-pattern analysis of LC-HRMS data to detect spills and compounds with high fluctuation in influent wastewater, J Hazard Mater, 361 (2019) 19-29.

[40] C. Hug, N. Ulrich, T. Schulze, W. Brack, M. Krauss, Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening, Environ Pollut, 184 (2014) 25-32.

[41] A.C. Chiaia-Hernandez, E.L. Schymanski, P. Kumar, H.P. Singer, J. Hollender, Suspect and nontarget screening approaches to identify organic contaminant records in lake sediments, Anal Bioanal Chem, 406 (2014) 7323-7335.

[42] M. Strynar, S. Dagnino, R. McMahen, S. Liang, A. Lindstrom, E. Andersen, L. McMillan, M. Thurman, I. Ferrer, C. Ball, Identification of Novel Perfluoroalkyl Ether Carboxylic Acids (PFECAs) and Sulfonic Acids (PFESAs) in Natural Waters Using Accurate Mass Time-of-Flight Mass Spectrometry (TOFMS), Environ Sci Technol, 49 (2015) 11622-11630.

[43] W. Brack, Effect-directed analysis: a promising tool for the identification of organic toxicants in complex mixtures?, Anal Bioanal Chem, 377 (2003) 397-407.

[44] C.M. Gallampois, E.L. Schymanski, M. Krauss, N. Ulrich, M. Bataineh, W. Brack, Multicriteria approach to select polyaromatic river mutagen candidates, Environ Sci Technol, 49 (2015) 2959-2968.

[45] Bruker Daltonics, Challenges in Metabolomics addressed by targeted and untargeted UHR-Q-TOF analysis https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/literature/ApplicationNotes/ET-21%20Metabolomics_ebook.pdf (Last accessed 15 February 2019).

[46] M. Meringer, E.L. Schymanski, Small Molecule Identification with MOLGEN and Mass Spectrometry, Metabolites, 3 (2013) 440-462.

[47] Royal Society of Chemistry, ChemSpider www.chemspider.com (Last accessed 15 February 2019).

[48] National Center for Biotechnology Information, Pubchem https://pubchem.ncbi.nlm.nih.gov (Last accessed 15 February 2019).

[49] M. Vinaixa, E.L. Schymanski, S. Neumann, M. Navarro, R.M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects, TrAC Trends Anal Chem, 78 (2016) 23-35.

[50] B. Petrie, R. Barden, B. Kasprzyk-Hordern, A review on emerging contaminants in wastewaters and the environment: current knowledge, understudied areas and recommendations for future monitoring, Water Res, 72 (2015) 3-27.

[51] M. Ibáñez, J.V. Sancho, F. Hernández, D. McMillan, R. Rao, Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-light mass spectrometry, TrAC Trends Anal Chem, 27 (2008) 481-489.

[52] E.L. Schymanski, H.P. Singer, P. Longree, M. Loos, M. Ruff, M.A. Stravs, C. Ripolles Vidal, J. Hollender, Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry, Environ Sci Technol, 48 (2014) 1811-1818.

[53] P. Gago-Ferrero, E.L. Schymanski, A.A. Bletsou, R. Aalizadeh, J. Hollender, N.S. Thomaidis, Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewaters with LC-HRMS/MS, Environ Sci Technol, 49 (2015) 12333-12341.

[54] P. Gago-Ferrero, E.L. Schymanski, J. Hollender, N.S. Thomaidis, Chapter 13 - Nontarget Analysis of Environmental Samples Based on Liquid Chromatography Coupled to High Resolution Mass Spectrometry (LC-HRMS), in: S. Perez, P. Eichhorn, D. Barcelo (Eds.) Applications of Time-of-Flight and Orbitrap Mass Spectrometry in Environmental, Food, Doping, and Forensic Analysis, Elsevier, 2016, pp. 381-403.

[55] Y. Verkh, M. Rozman, M. Petrovic, A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment, Chemosphere, 200 (2018) 397-404.

[56] R. Zaja, S. Terzic, I. Senta, J. Loncar, M. Popovic, M. Ahel, T. Smital, Identification of P-glycoprotein inhibitors in contaminated freshwater sediments, Environ Sci Technol, 47 (2013) 4813-4821.

[57] M.P. Schlusener, U. Kunkel, T.A. Ternes, Quaternary Triphenylphosphonium Compounds: A New Class of Environmental Pollutants, Environ Sci Technol, 49 (2015) 14282-14291.

[58] M.M. Plassmann, E. Tengstrand, K.M. Aberg, J.P. Benskin, Non-target time trend screening: a data reduction strategy for detecting emerging contaminants in biological samples, Anal Bioanal Chem, 408 (2016) 4203-4208.

[59] M. Ruff, M.S. Mueller, M. Loos, H.P. Singer, Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river

Rhine using high-resolution mass-spectrometry--Identification of unknown sources and compounds, Water Res, 87 (2015) 145-154.

[60] J.E. Schollee, E.L. Schymanski, S.E. Avak, M. Loos, J. Hollender, Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic, Anal Chem, 87 (2015) 12121-12129.

[61] J. von Neumann, Distribution of the Ratio of the Mean Square Succesive Difference to the Variance, The Annals of Mathematical Statistics, 4 (1941) 367-395.

[62] T. Bader, W. Schulz, K. Kummerer, R. Winzenbacher, General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry, Anal Chim Acta, 935 (2016) 173-186.

[63] Y.C. Tai, T.P. Speed, A multivariate empirical Bayes statistic for replicated microarray time course data, The Annals of Statistics, 34 (2006) 2387-2412.

[64] S. Kern, K. Fenner, H. Singer, R. Schwarzenbach, J. Hollender, Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry, Environ Sci Technol, 43 (2009) 7039-7046.

[65] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat Biotechnol, 30 (2012) 918-920.

[66] J.T. Prince, E.M. Marcotte, Chromatographic alignment of ESI-LC-MS Proteomics data sets by ordered bijective interpolated warping, Anal Chem, 45 (2006) 6140-6157.

[67] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Sluzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, Anal Chem, 78 (2006) 779-787.

[68] C. Kuhl, R. Tautenhahn, C. Bottcher, T.R. Larson, S. Neumann, CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets, Anal Chem, 84 (2012) 283-289.

[69] Bruker Daltonics, Challenges in Metabolomics addressed by targeted and untargeted UHR-Q-TOF analysis https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectrometry/Literature/literature/ApplicationNotes/ET-21%20Metabolomics_ebook.pdf (Last accessed 15 February 2019).

[70] S. Wolf, S. Schmidt, M. Muller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, BMC bioinformatics, 11 (2010) 148.

[71] M. Gerlich, S. Neumann, MetFusion: integration of compound identification strategies, J Mass Spectrom, 48 (2013) 291-298.

[72] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, MassBank: a public repository for sharing mass spectral data for life sciences, J Mass Spectrom, 45 (2010) 703-714.

[73] N.G. Dodder, K.M. Mullen, OrgMassSpecR: Organic Mass Spectrometry. R package version 0.4-4 (Last accessed 15 February 2019).

[74] E.S. Stein, Optimization and Testing of Mass Spectral Libray Search Algorithms for Compound Identification., J Am Soc Mass Spectrom, 5 (1994) 859-866.

[75] G. Libiseller, M. Dvorzak, U. Kleb, E. Gander, T. Eisenberg, F. Madeo, S. Neumann, G. Trausinger, F. Sinner, T. Pieber, C. Magnes, IPO: a tool for automated optimization of XCMS parameters, BMC Bioinformatics, 16 (2015) 118.

[76] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinformatics, 9 (2008) 504.

[77] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: a web-based platform to process untargeted metabolomic data, Anal Chem, 84 (2012) 5035-5039.

[78] M. Puig, L. Cabedo, J.J. Gracenea, A. Jiménez-Morales, J. Gámez-Pérez, J.J. Suay, Adhesion enhancement of powder coatings on galvanised steel by addition of organo-modified silica particles, Progress in Organic Coatings, 77 (2014) 1309-1315.

[79] J.A. Brown, M.D. M.P.H., Haz-Map®: Information on Hazardous Chemicals and Occupational Diseases, 2015.

[80] G.A. Burdock, Encyclopedia of food and color additives, in: G. Jaffe (Ed.) Encyclopedia of food and color additives, CRC Press, U.S.A., 1997, pp. 1751.

[81] E.L. Schymanski, H.P. Singer, J. Slobodnik, I.M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, N.S. Thomaidis, A. Bletsou, C. Zwiener, M. Ibanez, T. Portoles, R. de Boer, M.J. Reid, M. Onghena, U. Kunkel, W. Schulz, A. Guillon, N. Noyon, G. Leroy, P. Bados, S. Bogialli, D. Stipanicev, P. Rostkowski, J. Hollender, Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis, Anal Bioanal Chem, 407 (2015) 6237-6255.

[82] A. Kortenkamp, M. Faust, M. Scholze, T. Backhaus, Low-level exposure to multiple chemicals: reason for human health concerns?, Environ Health Perspect, 115 Suppl 1 (2007) 106-114.

[83] J.D. Pleil, Categorizing Biomarkers of the Human Exposome and Developing Metrics for Assessing Environmental Sustainability, J Toxicol Environ Health B Crit Rev, 15 (2012) 264-280.

[84] D.C.G. Muir, P.H. Howard, Are There Other Persistent Organic Pollutants? A Challenge for Environmental Chemists, Environ Sci Technol, 40 (2006) 7157-7166.

[85] J.E. Rager, M.J. Strynar, S. Liang, R.L. McMahen, A.M. Richard, C.M. Grulke, J.F. Wambaugh, K.K. Isaacs, R. Judson, A.J. Williams, J.R. Sobus, Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring, Environ Int, 88 (2016) 269-280.

[86] S.S. Andra, C. Austin, D. Patel, G. Dolios, M. Awawda, M. Arora, Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome, Environ Int, 100 (2017) 32-61.

[87] V. Leendert, H. Van Langenhove, K. Demeestere, Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments, TrAC Trends Anal Chem, 67 (2015) 192-208.

[88] M.M. Gomez-Ramos, C. Ferrer, O. Malato, A. Aguera, A.R. Fernandez-Alba, Liquid chromatography-high-resolution mass spectrometry for pesticide residue analysis in fruit and vegetables: screening and quantitative studies, J Chromatogr A, 1287 (2013) 24-37.

[89] L. Polgar, J.F. Garcia-Reyes, P. Fodor, A. Gyepes, M. Dernovics, L. Abranko, B. Gilbert-Lopez, A. Molina-Diaz, Retrospective screening of relevant pesticide metabolites in food using liquid chromatography high resolution mass spectrometry and accurate-mass databases of parent molecules and diagnostic fragment ions, J Chromatogr A, 1249 (2012) 83-91.

[90] A.C. Chiaia-Hernandez, M. Krauss, J. Hollender, Screening of lake sediments for emerging contaminants by liquid chromatography atmospheric pressure photoionization and electrospray ionization coupled to high resolution mass spectrometry, Environ Sci Technol, 47 (2013) 976-986.

[91] M. Gomez-Ramos Mdel, A. Perez-Parada, J.F. Garcia-Reyes, A.R. Fernandez-Alba, A. Aguera, Use of an accurate-mass database for the systematic identification of transformation products of organic contaminants in wastewater effluents, J Chromatogr A, 1218 (2011) 8002-8012.

[92] C. Moschet, I. Wittmer, J. Simovic, M. Junghans, A. Piazzoli, H. Singer, C. Stamm, C. Leu, J. Hollender, How a complete pesticide screening changes the assessment of surface water quality, Environ Sci Technol, 48 (2014) 5423-5432.

[93] P. Gago-Ferrero, V. Borova, M.E. Dasenaki, S. Tauhomaidis Nu, Simultaneous determination of 148 pharmaceuticals and illicit drugs in sewage sludge based on ultrasound-assisted extraction and liquid chromatography-tandem mass spectrometry, Anal Bioanal Chem, 407 (2015) 4287-4297.

[94] N.A. Alygizakis, P. Gago-Ferrero, V.L. Borova, A. Pavlidou, I. Hatzianestis, N.S. Thomaidis, Occurrence and spatial distribution of 158 pharmaceuticals, drugs of abuse and related metabolites in offshore seawater, Sci Total Environ, 541 (2016) 1097-1105.

[95] M. Zedda, C. Zwiener, Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools, Anal Bioanal Chem, 403 (2012) 2493-2502.

[96] A. Kaufmann, S. Walker, Evaluation of the interrelationship between mass resolving power and mass error tolerances for targeted bioanalysis using liquid chromatography coupled to high-resolution mass spectrometry, Rapid Commun Mass Spectrom, 27 (2013) 347-356.

[97] M. Mazzoni, M. Rusconi, S. Valsecchi, C.P. Martins, S. Polesello, An on-line solid phase extraction-liquid chromatography-tandem mass spectrometry method for the determination of perfluoroalkyl acids in drinking and surface waters, J Anal Methods Chem, 2015 (2015) 942016.

[98] D. Prats, F. Ruiz, B. Vazquez, R.-P. M., Removal of anionic and nonionic surfactants in a wastewater treatment plant with anaerobic digestion. A comparative study, Water Res, 31 (1997) 1925-1930.

[99] M.A. Aboulhassan, S. Souabi, A. Yaacoubi, M. Baudu, Removal of surfactant from industrial wastewaters by coagulation flocculation process, Int J Environ Sci Tech, 3 (2006) 327-332.

[100] S. Gonzalez, M. Petrovic, D. Barcelo, Removal of a broad range of surfactants from municipal wastewater--comparison between membrane bioreactor and conventional activated sludge treatment, Chemosphere, 67 (2007) 335-343.

[101] Y. Luo, W. Guo, H.H. Ngo, L.D. Nghiem, F.I. Hai, J. Zhang, S. Liang, X.C. Wang, A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment, Sci Total Environ, 473-474 (2014) 619-641.

[102] M. Jackson, C. Eadsforth, D. Schowanek, T. Delfosse, A. Riddle, N. Budgen, Comprehensive review of several surfactants in marine environments: Fate and ecotoxicity, Environ Toxicol Chem, 35 (2016) 1077-1086.

[103] K. Jardak, P. Drogui, R. Daghrir, Surfactants in aquatic and terrestrial environment: occurrence, behavior, and treatment processes, Environ Sci Pollut Res Int, 23 (2016) 3195-3216.

[104] G.-G. Ying, Fate, behavior and effects of surfactants and their degradation products in the environment, Environ Int, 32 (2006) 417-431.

[105] S. Samanipour, K. Langford, M.J. Reid, K.V. Thomas, A two stage algorithm for target and suspect analysis of produced water via gas chromatography coupled with high resolution time of flight mass spectrometry, J Chromatogr A, 1463 (2016) 153-161.

[106] S. Samanipour, J.A. Baz-Lomba, N.A. Alygizakis, M.J. Reid, N.S. Thomaidis, K.V. Thomas, Two stage algorithm vs commonly used approaches for the suspect screening of complex environmental samples analyzed via liquid chromatography high resolution time of flight mass spectroscopy: A test study, J Chromatogr A, 1501 (2017) 68-78.

[107] J. Acena, S. Stampachiacchiere, S. Perez, D. Barcelo, Advances in liquid chromatography-high-resolution mass spectrometry for quantitative and qualitative environmental analysis, Anal Bioanal Chem, 407 (2015) 6289-6299.

[108] E.L. Schymanski, A.J. Williams, Open Science for Identifying "Known Unknown" Chemicals, Environmental science & technology, 51 (2017) 5357-5359.

[109] Williams A., C.M. Grulke, McEachran A, A. Richard, Jolley R, Dunne J, Edmiston E, E. J, Comptox Chemistry Dashboard: Web-based data integration hub for environmental chemistry and toxicology data, https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=534970&Lab=NCCT (Last accessed 15 February 2019).

[110] V.G. Beretsou, A.K. Psoma, P. Gago-Ferrero, R. Aalizadeh, K. Fenner, N.S. Thomaidis, Identification of biotransformation products of citalopram formed in activated sludge, Water Res, 103 (2016) 205-214.

[111] M.C. Nika, A.A. Bletsou, E. Koumaki, C. Noutsopoulos, D. Mamais, A.S. Stasinakis, N.S. Thomaidis, Chlorination of benzothiazoles and benzotriazoles and transformation products identification by LC-HR-MS/MS, J Hazard Mater, 323 (2017) 400-413.

[112] C. Christophoridis, M.C. Nika, R. Aalizadeh, N.S. Thomaidis, Ozonation of ranitidine: Effect of experimental parameters and identification of transformation products, Sci Total Environ, 557-558 (2016) 170-182.

[113] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W.H. Tang, A. Römpp, S. Neumann, A.D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P.-A. Binzg, E.W. Deutsch, mzML—a Community Standard for Mass Spectrometry Data, Mol Cell Proteomics, 10 (2011).

[114] P.G. Pedrioli, J.K. Eng, R. Hubley, M. Vogelzang, E.W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R.H. Angeletti, R. Apweiler, K. Cheung, C.E. Costello, H. Hermjakob, S. Huang, R.K. Julian, E. Kapp, M.E. McComb, S.G. Oliver, G. Omenn, N.W. Paton, R. Simpson, R. Smith, C.F. Taylor, W. Zhu, R. Aebersold, A common open representation of mass spectrometry data and its application to proteomics research, Nat Biotechnol, 22 (2004) 1459-1466.

[115] B. Erickson, ANDI MS standard finalized, Anal Chem, 72 (2000) 103 A–103 A.

[116] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.T. Liu, M. Crusemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.C. Liaw, Y.L. Yang, H.U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A.B. P, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodriguez, A. Lamsa, C.

Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.M. Allard, P. Phapale, L.F. Nothias, T. Alexandrov, M. Litaudon, J.L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.O. Palsson, K. Pogliano, R.G. Linington, M. Gutierrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, Nat Biotechnol, 34 (2016) 828-837.

[117] M. Sturm, A. Bertsch, C. Gropl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, O. Kohlbacher, OpenMS - an open-source software framework for mass spectrometry, BMC Bioinformatics, 9 (2008) 163.

[118] K. Uppal, D.I. Walker, K. Liu, L. Shuzhao, Y.-M. G., J.D. P., Computational Metabolomics: A Framework for the Million Metabolome, Chem Res Toxicol, 29 (2016) 1956-1975.

[119] P.M. Allard, G. Genta-Jouve, J.L. Wolfender, Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification, Curr Opin Chem Biol, 36 (2017) 40-49.

[120] S. Samanipour, M.J. Reid, K.V. Thomas, Statistical Variable Selection: An Alternative Prioritization Strategy during the Nontarget Analysis of LC-HR-MS Data, Anal Chem, 89 (2017) 5585-5591.

[121] V. Dulio, B. van Bavel, E. Brorström-Lundén, J. Harmsen, J. Hollender, M. Schlabach, J. Slobodnik, K. Thomas, J. Koschorreck, Emerging pollutants in the EU: 10 years of NORMAN in support of environmental policies and regulations, Environ Sci Eur, 30 (2018).

[122] NORMAN SusDat, https://www.norman-network.com/nds/susdat/ (Last accessed 15 February 2019).

[123] NORMAN EMPODAT, Database of geo-referenced monitoring data on emerging substances in air, water and soil, https://www.norman-network.com/nds/empodat/ (Last accessed 15 February 2019).

[124] NORMAN ECOTOX, Platform for systematic collection and evaluation of the relevance and reliability of ecotoxicity studies, https://www.norman-network.com/nds/ecotox/ (Last accessed 15 February 2019).

[125] V. Dulio, J. Slobodnik, In response: the NORMAN perspectives on prioritization of emerging pollutants, Environ Toxicol Chem, 34 (2015) 2183-2185.

[126] B.Y.L. Peisl, E.L. Schymanski, P. Wilmes, Dark matter in host-microbiome metabolomics: Tackling the unknowns-A review, Anal Chim Acta, 1037 (2018) 13-27.

[127] K. Haug, R.M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. Gonzalez-Beltran, S.A. Sansone, J.L. Griffin, C. Steinbeck, MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data, Nucleic Acids Res, 41 (2013) D781-786.

[128] T. Schulze, E.L. Schymanski, M. Stravs, S. Neumann, M. Krauss, H. Singer, C. Hug, C. Gallampois, J. Hollender, J. Slobodnik, B. Werner, NORMAN MassBank. Towards a community-driven, open-access accurate mass spectral database for the identification of emerging pollutants, NORMAN Net Bull, 3 (2012) 9-10.

[129] NORMAN MassBank, NORMAN MassBank, Database of mass spectra of unknown or provisionally identified substances, www.massbank.eu (Last accessed 15 February 2019).

[130] C. Frainay, E.L. Schymanski, S. Neumann, B. Merlet, R.M. Salek, F. Jourdan, O. Yanes, Mind the Gap: Mapping Mass Spectral Databases in Genome-Scale Metabolic Networks Reveals Poorly Covered Areas, Metabolites, 8 (2018).

[131] NormaNEWS, The Norman Early Warning System - a new retrospective analysis exchange platform for laboratories active in non-target screening (http://www.normandata.eu/?q=node/244 and https://comptox.epa.gov/dashboard/chemical_lists/normanews (Last accessed 15 February 2018).

[132] EMBLAS Project, Improving environmental monitoring in the Black Sea - Phase II (EMBLAS-II) ENPI/2013/313-169, UNDP Regional Bureau for Europe and the CIS. In: Slobodnik J et al. (eds). http://emblasproject.org/wp-content/uploads/2017/04/Joint-Black-Sea-Surveys-2016_16.pdf (Last accessed 15 February 2019).

[133] G. McEneff, L. Barron, B. Kelleher, B. Paull, B. Quinn, A year-long study of the spatial occurrence and relative distribution of pharmaceutical residues in sewage effluent, receiving marine waters and marine bivalves, Sci Total Environ, 476-477 (2014) 317-326.

[134] B. Berglund, Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics, Infect Ecol Epidemiol, 5 (2015) 28564.

[135] World Health Organization, Antibiotic resistance - a threat to global health security, http://www.who.int/drugresistance/activities/wha66_side_event/en/ (Last accessed 15 February 2019).

[136] P. Gago-Ferrero, V. Borova, M.E. Dasenaki, N.S. Thomaidis, Simultaneous determination of 148 pharmaceuticals and illicit drugs in sewage sludge based on ultrasound-assisted extraction and liquid chromatography-tandem mass spectrometry, Anal Bioanal Chem, 407 (2015) 4287-4297.

[137] NORMAN Suspect List Exchange, https://www.norman-network.com/?q=node/236 (Last accessed 15 February 2019).

[138] HighChem Ltd, mzCloud mass spectral database https://mzcloud.org/ (Last accessed 15 February 2019).

[139] M. Katajamaa, J. Miettinen, M. Oresic, MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, Bioinformatics, 22 (2006) 634-636.

[140] M. Loos, nontarget: Detecting isotope, adduct and homologue relations in LC-MS data. R package version 1.9. https://CRAN.R-project.org/package=nontarget (Last accessed 15 February 2019).

[141] C.D. Broeckling, F.A. Afsar, S. Neumann, A. Ben-Hur, J.E. Prenni, RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data, Anal Chem, 86 (2014) 6812-6817.

[142] Aalizadeh et al., Development and Prediction of Liquid Chromatographic Retention Time Indices. in preparation.

[143] International Commission for the Protection of the Danube River (ICPDR), Joint Danube Survey 3. A comprehensive analysis of Danube water quality (http://www.danubesurvey.org/jds-files/nodes/documents/jds3_final_scientific_report_1.pdf), Vienna, Austria, 2015.

[144] P. Gago-Ferrero, A. Bletsou, D.E. Damalas, R. Aalizadeh, A.N. Alygizakis, H. Singer, J. Hollender, N.S. Thomaidis, Wide-scope quantitative target screening of 2248 emerging contaminants in wastewater samples with UPLC-Q-ToF-HRMS/MS, in preparation.

[145] S. Yan, L. Chen, X. Dou, M. Qi, Q. Du, Q. He, M. Nan, Z. Chang, P. Nan, Toxicity of 8-Hydroxyquinoline in Cryprinus carpio Using the Acute Toxicity Test, Hepatase Activity Analysis and the Comet Assay, Bull Environ Contam Toxicol, 95 (2015) 171-176.

[146] I. Liska, The Danube River Basin, Springer-Verlag, Berlin, Heidelberg, 2015.

[147] S. Terzic, I. Senta, M. Ahel, M. Gros, M. Petrovic, D. Barcelo, J. Muller, T. Knepper, I. Marti, F. Ventura, P. Jovancic, D. Jabucar, Occurrence and fate of emerging wastewater contaminants in Western Balkan Region, Sci Total Environ, 399 (2008) 66-77.

[148] F. Heeb, H. Singer, B. Pernet-Coudrier, W. Qi, H. Liu, P. Longree, B. Muller, M. Berg, Organic micropollutants in rivers downstream of the megacity Beijing: sources and mass fluxes in a large-scale wastewater irrigation system, Environ Sci Technol, 46 (2012) 8680-8688.

[149] M. Konig, B.I. Escher, P.A. Neale, M. Krauss, K. Hilscherova, J. Novak, I. Teodorovic, T. Schulze, S. Seidensticker, M.A. Kamal Hashmi, J. Ahlheim, W. Brack, Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis, Environ Pollut, 220 (2017) 1220-1230.

[150] T. Mackulak, M. Mosny, J. Skubak, R. Grabic, L. Birosova, Fate of psychoactive compounds in wastewater treatment plant and the possibility of their degradation using aquatic plants, Environ Toxicol Pharmacol, 39 (2015) 969-973.

[151] T. Kosjek, S. Perko, M. Zupanc, M. Zanoski Hren, T. Landeka Dragicevic, D. Zigon, B. Kompare, E. Heath, Environmental occurrence, fate and transformation of benzodiazepines in water treatment, Water Res, 46 (2012) 355-368.

[152] J. Fick, T. Brodin, M. Heynen, J. Klaminder, M. Jonsson, K. Grabicova, T. Randak, R. Grabic, V. Kodes, J. Slobodnik, A. Sweetman, M. Earnshaw, A. Barra Caracciolo, T. Lettieri, R. Loos, Screening of benzodiazepines in thirty European rivers, Chemosphere, 176 (2017) 324-332.

[153] I. Krizman-Matasic, P. Kostanjevecki, M. Ahel, S. Terzic, Simultaneous analysis of opioid analgesics and their metabolites in municipal wastewaters and river water by liquid chromatography-tandem mass spectrometry, J Chromatogr A, 1533 (2018) 102-111.

[154] M. Clara, O. Gans, S. Weiss, D. Sanz-Escribano, S. Scharf, C. Scheffknecht, Perfluorinated alkylated substances in the aquatic environment: an Austrian case study, Water Res, 43 (2009) 4760-4768.

[155] T. Smital, S. Terzic, R. Zaja, I. Senta, B. Pivcevic, M. Popovic, I. Mikac, K.E. Tollefsen, K.V. Thomas, M. Ahel, Assessment of toxicological profiles of the municipal wastewater effluents using chemical analyses and bioassays, Ecotoxicol Environ Saf, 74 (2011) 844-851.

[156] I. Lupan, R. Carpa, A. Oltean, B.S. Kelemen, O. Popescu, Release of Antibiotic Resistant Bacteria by a Waste Treatment Plant from Romania, Microbes Environ, 32 (2017) 219-225.

[157] L. Birosova, T. Mackulak, I. Bodik, J. Ryba, J. Skubak, R. Grabic, Pilot study of seasonal occurrence and distribution of antibiotics and drug resistant bacteria in wastewater treatment plants in Slovakia, Sci Total Environ, 490 (2014) 440-444.

[158] D. Fatta-Kassinos, C. Manaia, T.U. Berendonk, E. Cytryn, J. Bayona, B. Chefetz, J. Slobodnik, N. Kreuzinger, L. Rizzo, S. Malato, L. Lundy, A. Ledin, COST Action ES1403: new and emerging challenges and opportunities in wastewater reuse (NEREUS), Environ Sci Pollut Res Int, 22 (2015) 7183-7186.

[159] W. Brack, R. Altenburger, G. Schuurmann, M. Krauss, D. Lopez Herraez, J. van Gils, J. Slobodnik, J. Munthe, B.M. Gawlik, A. van Wezel, M. Schriks, J. Hollender, K.E. Tollefsen, O. Mekenyan, S. Dimitrov, D. Bunke, I. Cousins, L. Posthuma, P.J. van den Brink, M. Lopez de Alda, D. Barcelo, M. Faust, A. Kortenkamp, M. Scrimshaw, S. Ignatova, G. Engelen, G. Massmann, G. Lemkine, I. Teodorovic, K.H. Walz, V. Dulio, M.T. Jonker, F. Jager, K. Chipman, F. Falciani, I. Liska, D. Rooke, X. Zhang, H. Hollert, B. Vrana, K. Hilscherova, K. Kramer, S. Neumann, R. Hammerbacher, T. Backhaus, J. Mack, H. Segner, B. Escher, G. de Aragao Umbuzeiro, The SOLUTIONS project: challenges and responses for present and future emerging pollutants in land and water resources management, Sci Total Environ, 503-504 (2015) 22-31.

[160] V. Dulio, P.C. von der Ohe, NORMAN prioritisation framework for emerging substances, http://www.norman-network.net/sites/default/files/norman_prioritisation_manual_15%20April2013_final_for_website.pdf, (Last accessed 15 February 2019).

[161] P.C. von der Ohe, V. Dulio, J. Slobodnik, E. De Deckere, R. Kuhne, R.U. Ebert, A. Ginebreda, W. De Cooman, G. Schuurmann, W. Brack, A new risk assessment approach for the prioritization of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the

European Water Framework Directive, Sci Total Environ, 409 (2011) 2064-2077.

[162] N.S. Thomaidis, P. Gago-Ferrero, C. Ort, N.C. Maragou, N.A. Alygizakis, V.L. Borova, M.E. Dasenaki, Reflection of Socioeconomic Changes in Wastewater: Licit and Illicit Drug Use Patterns, Environ Sci Technol, 50 (2016) 10065-10072.

[163] R. van der Oost, G. Sileno, T. Janse, M.T. Nguyen, H. Besselink, A. Brouwer, SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part II-field feasibility survey, Environ Toxicol Chem, 36 (2017) 2400-2416.

[164] B.I. Escher, U.-A.y.U.S. Asmall yi, P.A. Behnisch, W. Brack, F. Brion, A. Brouwer, S. Buchinger, S.E. Crawford, D. Du Pasquier, T. Hamers, K. Hettwer, K. Hilscherova, H. Hollert, R. Kase, C. Kienle, A.J. Tindall, J. Tuerk, R. van der Oost, E. Vermeirssen, P.A. Neale, Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive, Sci Total Environ, 628-629 (2018) 748-765.

[165] M.R. Gillings, W.H. Gaze, A. Pruden, K. Smalla, J.M. Tiedje, Y.G. Zhu, Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution, ISME J, 9 (2015) 1269-1279.

[166] F. Been, L. Rossi, C. Ort, S. Rudaz, O. Delémont, P. Esseiva, Population normalization with ammonium in wastewater-based epidemiology: application to illicit drug monitoring., Environ Sci Technol, 48 (2014) 8162-8169.

[167] J. Slobodnik, L. Mrafkova, M. Carere, F. Ferrara, B. Pennelli, G. Schrüümann, P.C. Von der Ohe, Identification of river basin specific pollutants and derivation of environmental quality standards: A case study in the Slovak Republic, TrAC Trends Anal Chem, 41 (2012) 133-145.

[168] E. Marti, J. Jofre, J.L. Balcazar, Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant, PLoS One, 8 (2013) e78906.

[169] B. Berglund, J. Fick, P.E. Lindgren, Urban wastewater effluent increases antibiotic resistance gene concentrations in a receiving northern European river, Environ Toxicol Chem, 34 (2015) 192-196.

[170] S. Rodriguez-Mozaz, S. Chamorro, E. Marti, B. Huerta, M. Gros, A. Sanchez-Melsio, C.M. Borrego, D. Barcelo, J.L. Balcazar, Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river, Water Res, 69 (2015) 234-242.

[171] L. Proia, D. von Schiller, A. Sanchez-Melsio, S. Sabater, C.M. Borrego, S. Rodriguez-Mozaz, J.L. Balcazar, Occurrence and persistence of antibiotic resistance genes in river biofilms after wastewater inputs in small rivers, Environ Pollut, 210 (2016) 121-128.

[172] N. Makowska, R. Koczura, J. Mokracka, Class 1 integrase, sulfonamide and tetracycline resistance genes in wastewater treatment plant and surface water, Chemosphere, 144 (2016) 1665-1673.

[173] R. Loos, G. Locoro, S. Contini, Occurrence of polar organic contaminants in the dissolved water phase of the Danube River and its major tributaries using SPE-LC-MS(2) analysis, Water Res, 44 (2010) 2325-2335.

[174] R. Loos, S. Tavazzi, G. Mariani, G. Suurkuusk, B. Paracchini, G. Umlauf, Analysis of emerging organic contaminants in water, fish and suspended particulate matter (SPM) in the Joint Danube Survey using solid-phase extraction followed by UHPLC-MS-MS and GC-MS analysis, Sci Total Environ, 607-608 (2017) 1201-1212.

[175] Z. Tousova, P. Oswald, J. Slobodnik, L. Blaha, M. Muz, M. Hu, W. Brack, M. Krauss, C. Di Paolo, Z. Tarcai, T.B. Seiler, H. Hollert, S. Koprivica, M. Ahel, J.E. Schollee, J. Hollender, M.J. Suter, A.O. Hidasi, K. Schirmer, M. Sonavane, S. Ait-Aissa, N. Creusot, F. Brion, J. Froment, A.C. Almeida, K. Thomas, K.E. Tollefsen, S. Tufi, X. Ouyang, P. Leonards, M. Lamoree, V.O. Torrens, A. Kolkman, M. Schriks, P. Spirhanzlova, A. Tindall, T. Schulze, European demonstration program on the effect-based and chemical identification and monitoring of organic pollutants in European surface waters, Sci Total Environ, 601-602 (2017) 1849-1868.

[176] M. Biel-Maeso, R.M. Baena-Nogueras, C. Corada-Fernandez, P.A. Lara-Martin, Occurrence, distribution and environmental risk of pharmaceutically active compounds (PhACs) in coastal and ocean waters from the Gulf of Cadiz (SW Spain), Sci Total Environ, 612 (2018) 649-659.

[177] E. Castrignano, A.M. Kannan, E.J. Feil, B. Kasprzyk-Hordern, Enantioselective fractionation of fluoroquinolones in the aqueous environment using chiral liquid chromatography coupled with tandem mass spectrometry, Chemosphere, 206 (2018) 376-386.

[178] A. Diuzheva, J. Balogh, J. Jeko, Z. Cziaky, Application of liquid-liquid microextraction for the effective separation and simultaneous determination of 11 pharmaceuticals in wastewater samples using high-performance liquid chromatography with tandem mass spectrometry, J Sep Sci, DOI 10.1002/jssc.201800309 (2018) 1-8.

[179] M. Farre, M. Petrovic, M. Gros, T. Kosjek, E. Martinez, E. Heath, P. Osvald, R. Loos, K. Le Menach, H. Budzinski, F. De Alencastro, J. Muller, T. Knepper, G. Fink, T.A. Ternes, E. Zuccato, P. Kormali, O. Gans, R. Rodil, J.B. Quintana, F. Pastori, A. Gentili, D. Barcelo, First interlaboratory exercise on non-steroidal anti-inflammatory drugs analysis in environmental samples, Talanta, 76 (2008) 580-590.

[180] Z. Moldovan, R. Chira, A.C. Alder, Environmental exposure of pharmaceuticals and musk fragrances in the Somes River before and after upgrading the municipal wastewater treatment plant Cluj-Napoca, Romania, Environ Sci Pollut Res Int, 16 Suppl 1 (2009) S46-54.

[181] Z. Moldovan, G. Schmutzer, F. Tusa, R. Calin, A.C. Alder, An overview of pharmaceuticals and personal care products contamination along the river Somes watershed, Romania, J Environ Monit, 9 (2007) 986-993.

[182] B. Fabry, Tenside. Eigenschaften, Rohstoffe, Produktion, Anwendungen. Chem. Unserer Zeit 25, 214–222, (1991).

[183] M.J. Rosen, J.T. Kunjappu, Surfactants and Interfacial Phenomena, Wiley, 2012.

[184] CESIO, Comité Européen des Agents de Surface et de Leurs Intermédiaires Organiques (CESIO), Surfactants Production EU 1994 - 2016 http://www.cesio.eu/images/content/CESIO-Statistics-2016-for-website.pdf (Last accessed 15 February 2019).

[185] Credence Research, Global Surfactants Market Is Expected To Reach US$ 47.16 Bn By 2023: Rising Demand From End Use Applications Will Fuel The Market Growth. Available at: https://www.credenceresearch.com/press/global-surfactants-market (Last accessed 15 February 2019).

[186] UNEP, OECD SIDS Linear alkylbenzene sulfonate (LAS) Initial assessment report for 20th SIAM, Paris, France, 19-21 April, 2005 https://hpvchemicals.oecd.org/ui/handler.axd?id=5b837fb0-350c-4742-914e-5f6513df120a (Last accessed 15 February 2019).

[187] HERA, Human & Environmental Risk Assessment (HERA) on ingredients of European household cleaning products. Linear Alkylbenzene Sulphonate. Revised Environmental Aspect of the HERA Report, 2013.

[188] G.G. Ying, Fate, behavior and effects of surfactants and their degradation products in the environment, Environ Int, 32 (2006) 417-431.

[189] HERA, Human & Environmental Risk Assessment (HERA) on ingredients of European household cleaning products. Alcohol Ethoxysulphates (AES) Environmental Risk Assessment., 2004.

[190] K.A. Massey, A.M. Snelling, A. Nicolaou, Quantitative analysis of surfactant deposits on human skin by liquid chromatography/electrospray ionisation tandem mass spectrometry, Rapid Commun Mass Spectrom, 24 (2010) 1371-1376.

[191] D.C. McAvoy, D.D. Scott, N.j. Fendinger, W.S. Eckhoff, Removal of alcohol ethoxylates, alkyl ethoxylate sulfates, and linear alkylbenzene sulfonates in wastewater treatment, Environ Toxicol Chem, 17 (1998) 1705-1711.

[192] P.A. Lara-Martin, E. Gonzalez-Mazo, B.J. Brownawell, Environmental analysis of alcohol ethoxylates and nonylphenol ethoxylate metabolites by ultra-performance liquid chromatography-tandem mass spectrometry, Anal. Bioanal. Chem., 402 (2012) 2359-2368.

[193] Y.-L. Huang, Q.-B. Li, X. Deng, Y.-H. Lu, X.-K. Liao, M.-Y. Hong, Y. Wang, Aerobic and anaerobic biodegradation of polyethylene glycols using sludge microbes, Process Biochemistry, 40 (2005) 207-211.

[194] J.D. Rogers, E.M. Thurman, I. Ferrer, J.S. Rosenblum, M.V. Evans, P.J. Mouser, J.N. Ryan, Degradation of polyethylene glycols and polypropylene glycols in microcosms simulating a spill of produced water in shallow groundwater, Environ Sci Process Impacts, DOI 10.1039/c8em00291f (2018).

[195] A. Di Corcia, F. Casassa, C. Crescenzi, A. Marcomini, R. Samperi, Investigation of the fate of linear alkyl benzenesulfonates and coproducts in a

laboratory biodgradation test by using Liquid Chromatography/Mass Spectrometry, Environ Sci Technol, 33 (1999) 4112-4118.

[196] P.H. Brunner, S. Capri, A. Marcomini, W. Giger, Occurrence and behaviour of linear alkylbenzenesulphonates, nonylphenol, nonylphenol mono- and nonylphenol diethoxylates in sewage and sewage sludge treatment, Water Res, 22 (1988) 1465-1472.

[197] F.R. Schröder, M. Schmitt, U. Reichensperger, Effect of wastew water treatment technology on the elimination of anionic surfactants, Waste Manag, 19 (1999) 125-131.

[198] D.R. Helsel, More than obvious: Better methods for interpreting nondetect data, Environ Sci Technol, 39 (2005) 419A-423A.

[199] D.R. Helsel, Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it, Chemosphere, 65 (2006) 2434-2439.

[200] N.A. Alygizakis, P. Oswald, N.S. Thomaidis, E.L. Schymanski, R. Aalizadeh, T. Schulze, M. Oswaldova, J. Slobodnik, NORMAN Digital Sample Freezing Platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in “digitally frozen” environmental samples,Trends Anal Chem, Accepted.

[201] C. Corada-Fernandez, P.A. Lara-Martin, L. Candela, E. Gonzalez-Mazo, Tracking sewage derived contamination in riverine settings by analysis of synthetic surfactants, J Environ Monit, 13 (2011) 2010-2017.

[202] A. Di Corcia, C. Crescenzi, A. Marcomini, R. Samperi, Liquid chromatography-Electrospray-Mass Spectrometry as a valuable tool for characterizing biodegradation intermediates of branched alcohol ethoxylate surfactants, Environ Sci Technol, 32 (1998) 711-718.

[203] A. Di Corcia, L. Capuani, F. Casassa, A. Marcomini, R. Samperi, Fate of linear alkyl benzenesulfonates, coproducts and their metabolites in sewage treatment plants and in receiving river waters, Environ Sci Technol, 33 (1999) 4119-4125.

[204] J.A. Field, T.M. Field, T. Poiger, W. Giger, Determination of secondary alkane sulfonates in sewage wastewaters by solid-phase extraction and injection-port derivatization Gas Chromatography/Mass Spectrometry, Environ Sci Pollut Res Int, 28 (1994) 497-503.

[205] M. Gonsior, M. Zwartjes, W.J. Cooper, W. Song, K.P. Ishida, L.Y. Tseng, M.K. Jeung, D. Rosso, N. Hertkorn, P. Schmitt-Kopplin, Molecular characterization of effluent organic matter identified by ultrahigh resolution mass spectrometry, Water Res, 45 (2011) 2943-2953.

[206] S. Gonzalez, M. Petrovic, M. Radetic, P. Jovancic, V. Ilic, D. Barcelo, Characterization and quantitative analysis of surfactants in textile wastewater by liquid chromatography/quadrupole-time-of-flight mass spectrometry, Rapid Commun Mass Spectrom, 22 (2008) 1445-1454.

[207] S. Riediker, M.J.-F. Suter, W. Giger, Benzene- and naphthalenesulfonates in leachates and plumes of landfills, Water Res, 34 (2000) 2069-2079.

[208] J.M. Traverso-Soto, P.A. Lara-Martin, E. Gonzalez-Mazo, V.M. Leon, Distribution of anionic and nonionic surfactants in a sewage-impacted Mediterranean coastal lagoon: inputs and seasonal variations, Sci Total Environ, 503-504 (2015) 87-96.

[209] S. Cantarero, A. Zafra-Gomez, O. Ballesteros, A. Navalon, J.L. Vilchez, C. Verge, J.A. De Ferrer, Matrix effect study in the determination of linear alkylbenzene sulfonates in sewage sludge samples, Environ Toxicol Chem, 30 (2011) 813-818.

[210] J.J. Gray, L.K. Kanagy, E.T. Furlong, J.W. McCoy, C.J. Kanagy, Determination of the anionic surfactant di(ethylhexyl) sodium sulfosuccinate in water samples collected from Gulf of Mexico coastal waters before and after landfall of oil from the Deepwater Horizon oil spill, May to October, 2010. U.S. Geological Survey Open-File Report 2010–1318, 15 p. (Available at http://pubs.usgs.gov/of/2010/1318/, Last accessed 15 February 2019).

[211] T.P. Knepper, D. Barceló, P. De Voogt, Analysis and fate of surfactants in the aquatic environment. 1st. Amsterdam, Boston: Elsevier (Wilson & Wilson's comprehensive analytical chemistry, v. 40), 2003.

[212] N. Sakai, J. Shirasaka, Y. Matsui, M.R. Ramli, K. Yoshida, M. Ali Mohd, M. Yoneda, Occurrence, fate and environmental risk of linear alkylbenzene sulfonate in the Langat and Selangor River basins, Malaysia, Chemosphere, 172 (2017) 234-241.

[213] K. McDonough, K. Casteel, N. Itrich, J. Menzies, S. Belanger, K. Wehmeyer, T. Federle, Evaluation of anionic surfactant concentrations in US effluents and probabilistic determination of their combined ecological risk in mixing zones, Sci Total Environ, 572 (2016) 434-441.

[214] P.A. Lara-Martin, A. Gomez-Parra, E. Gonzalez-Mazo, Sources, transport and reactivity of anionic and non-ionic surfactants in several aquatic ecosystems in SW Spain: a comparative study, Environmental pollution, 156 (2008) 36-45.

[215] E. Matthijs, M.S. Holt, A. Kiewiet, G.B.J. Rijs, Environmental monitoring for linear alkylbenzene sulfonate, alcohol ethoxylate, alchohol ethoxy sulfate, alcohol sulfate, and soap, Environ Toxicol Chem, 18 (1999) 2634-2644.

[216] L. Lunar, S. Rubio, D. Perez-Bendito, Analysis of linear alkylbenzene sulfonate homologues in environmental water samples by mixed admicelle-based extraction and liquid chromatography/mass spectrometry, Analyst, 131 (2006) 835-841.

[217] J. Riu, P. Eichhorn, J.A. Guerrero, T.P. Knepper, D. Barceló, Determination of linear alkylbenzenesulfonates in wastewater treatment plants and coastal waters by automated solid-phase extraction followed by capillary electrophoresis–UV detection and confirmation by capillary electrophoresis–mass spectrometry, J Chromatogr A, 889 (2000) 221-229.

[218] M. Holt, K. Fox, M. Daniel, H. Buckland, Linear alkylbenzene sulfonate and boron monitoring in four catchments in the UK contribution to GREAT-ER #11, Sci Total Environ, 314-316 (2003) 271-288.

[219] M. Clara, S. Scharf, C. Scheffknecht, O. Gans, Occurrence of selected surfactants in untreated and treated sewage, Water Res, 41 (2007) 4339-4348.

[220] C. Sparham, N. Rehman, J. Melling, J. Van Duynhoven, S. Marshall, Biodegradability of highly ethoxylated nonionic surfactants: Determination of intermediates and pathways of biodegradation, Environ Toxicol Chem, 27 (2008) 1069-1067.

[221] J.M. Traverso-Soto, P.A. Lara-Martin, V.M. Leon, E. Gonzalez-Mazo, Analysis of alcohol polyethoxylates and polyethylene glycols in marine sediments, Talanta, 110 (2013) 171-179.

[222] M. Castillo, E. Martínez, D. Barceló, A. Ginebreda, L. Tirapu, Determination of non-ionic surfactants and polar degradation products in influent and effluent water samples and sludges of sewage treatment plants by a generic solid-phase extraction protocol, The Analyst, 125 (2000) 1733-1739.

[223] A. Marcomini, M. Zanette, G. Pojana, M.J.-F. Suter, Behavior of aliphatic alcohol polyethoxylates and their metabolites under standardized aerobic biodegradation conditions, Environ Toxicol Chem, 19 (2000) 549-554.