

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

4 Nikiforos A. Alygizakis^{1,2†}, Saer Samanipour^{3†}, Juliane Hollender^{4,5}, María Ibáñez⁶, Sarit Kaserzon⁷, Varvara
5 Kokkali⁸, Jan A. van Leerdam⁹, Jochen F. Mueller⁷, Martijn Pijnappels¹⁰, Malcolm J. Reid³, Emma L.
6 Schymanski^{4,11}, Jaroslav Slobodnik², Nikolaos S. Thomaidis¹, Kevin V. Thomas^{3,7*}

7
8 ¹Laboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Panepistimiopolis
9 Zografou, 15771 Athens, Greece

10 ²Environmental Institute, s.r.o., Okružná 784/42, 972 41 Koš, Slovak Republic

11 ³Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway

12 ⁴Eawag: Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

13 ⁵Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092 Zürich, Switzerland

14 ⁶Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, 12071 Castellón de
15 la Plana, Spain

16 ⁷Queensland Alliance for Environmental Health Sciences (QAEHS), The University of Queensland, 20
17 Cornwall Street, Woolloongabba, Queensland, 4102 Australia

18 ⁸Vitens Laboratory, Snekerterweg 61, 8912 AA Leeuwarden, The Netherlands

19 ⁹KWR Watercycle Research Institute, P.O. Box 1072, 3430 BB, Nieuwegein, The Netherlands

20 ¹⁰Rijkswaterstaat, Ministry of Infrastructure and the Environment, Zuiderwagenplein 2, 8224 AD, Lelystad,
21 The Netherlands

22 ¹¹Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 7, Avenue des Hauts
23 Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg

24
25 [†]Authors contributed equally.

26 *Corresponding author

27 Kevin V Thomas

28 Queensland Alliance for Environmental Health Sciences (QAEHS), The University of Queensland, 20
29 Cornwall Street, Woolloongabba, Queensland, 4102 Australia.

30 Email: kevin.thomas@uq.edu.au

31 Phone: 0061 417287582

32 **Manuscript details**

33 Word count abstract: 225

34 Word count text: 6735

35 Keywords: suspect screening, high-resolution mass spectrometry, retrospective screening, early warning
36 system, contaminants of emerging concern

37

38 Abstract

39 A key challenge in the environmental and exposure sciences is to establish experimental evidence of the
40 role of chemical exposure in human and environmental systems. High resolution and accurate tandem mass
41 spectrometry (HRMS) is increasingly being used for the analysis of environmental samples. One lauded
42 benefit of HRMS is the possibility to retrospectively process data for (previously omitted) compounds that
43 has led to the archiving of HRMS data. Archived HRMS data affords the possibility of exploiting historical
44 data to rapidly and effectively establish the temporal and spatial occurrence of newly identified
45 contaminants through retrospective suspect screening. We propose to establish a global emerging
46 contaminant early warning network to rapidly assess the spatial and temporal distribution of contaminants
47 of emerging concern in environmental samples through performing retrospective analysis on HRMS data.
48 The effectiveness of such a network is demonstrated through a pilot study, where eight reference
49 laboratories with available archived HRMS data retrospectively screened data acquired from aqueous
50 environmental samples collected in 14 countries on 3 different continents. The widespread spatial
51 occurrence of several surfactants (e.g. [PEGs](#) and [C12AEO-PEGs](#)), transformation products of selected drugs
52 (e.g. gabapentin-lactam, metoprolol-acid, carbamazepine-10-hydroxy, omeprazole-4-hydroxy-sulphide, 2-
53 benzothiazole-sulfonic-acid), and industrial chemicals (3-nitrobenzenesulfonate and bisphenol-S) was
54 revealed. Obtaining identifications of increased reliability through retrospective suspect screening is
55 challenging and recommendations for dealing with issues such as broad chromatographic peaks, data
56 acquisition, and sensitivity are provided.

57

58 Introduction

59 One of the key challenges in the environmental and exposure sciences is to establish experimental evidence
60 of the role of chemical exposure in human and environmental systems.^{1,2} Our ‘chemosphere’ is
61 continuously changing and most chemicals that are indexed in the Chemical Abstract Service (CAS) are not
62 characterized with respect to their potential effects on human safety and environmental health.³ Non-
63 target analysis employing high-resolution mass spectrometers has been established over the past years as
64 one of the key approaches for tackling this complexity. High resolution and accurate hybrid tandem mass
65 spectrometers, such as time-of-flight and Orbitrap instruments have facilitated increased reliability in
66 target analysis (using reference standards), enabled suspect screening (without reference standards) and
67 screening for unknowns.⁴⁻⁶ Substantial research effort has been placed on developing tools and workflows
68 that expedite these three approaches, with the overall outcome that the contemporary analyst is able to
69 obtain large amount of accurate mass data for a particular sample. For example, in 2013 the NORMAN
70 Network of reference laboratories, research centres and related organisations for monitoring of emerging
71 environmental substances (www.norman-network.net) organized a non-target screening collaborative trial
72 employing target, suspect, and non-target workflows to identify substances in water samples.⁷ This trial
73 revealed that non-target techniques are in general substantially harmonized between practitioners and
74 that although data processing can be time consuming and remains a major bottleneck, suspect screening
75 approaches are very popular. However it recognized that *“better integration and connection of desired
76 features into software packages, the exchange of target and suspect lists, and the contribution of more
77 spectra from standard substances into (openly accessible) database”* are necessary for the technique to

78 reach maturity.⁴ The archiving of HRMS data also allows for data to be processed retrospectively, for
79 example to investigate the occurrence of a newly identified compound or simply one that was not
80 considered at the time of analysis.⁸ This possibility has led to researchers working in this field to digitally
81 archive data in preparation for future retrospective analysis and has even led to proposals for the
82 establishment of data repositories, akin to environmental data banks, where digital information can be
83 safely stored for future retrospective analysis.

84 Non-target HRMS full scan data allows the potential for rapid and cost-effective screening of the occurrence
85 of newly identified contaminants in previously archived HRMS data; often referred to as retrospective
86 analysis. Typically, it refers to the application of suspect screening workflows to archived data as reference
87 standard measurements are not available for the analytical settings. Whilst retrospective analysis with
88 HRMS in environmental sciences has been discussed for some time^{7,8,9,10} there are few published studies
89 that actually apply the approach^{11,12}. As far as we are aware there have not been coordinated studies to
90 investigate the spatial and temporal distribution of contaminants of emerging concern in environmental
91 samples through performing retrospective analysis on HRMS data acquired using different instrumental
92 platforms and data processing software. This has the potential to be an improved and effective strategy for
93 establishing the extent of a newly identified contaminant's occurrence rather than the traditional approach
94 of a new contaminant(s) being reported in the scientific literature and individual research groups
95 subsequently validating targeted methods and reporting their own data. In order to test this hypothesis, a
96 pilot study was performed where eight reference laboratories with available archived HRMS data were
97 recruited with the goal of exploring the potential of a contaminant of emerging concern early warning
98 network through the use of retrospective suspect screening employing HRMS. The pilot study was referred
99 to as the NORMAN Early Warning System, abbreviated to NormaNEWS.¹³

100

101 **Materials and Methods**

102 **Participants and samples**

103 The participants of the NormaNEWS exercise (8 reference laboratories; Eawag, KWR, NIVA, QAEHS, RWS,
104 UJI, UoA, and Vitens) submitted samples from 14 countries and 3 continents. In total 48 sets of data from
105 the analysis of environmental samples were evaluated. Detailed information on sample matrix, sampling
106 date, instrument type, chromatographic separation (flow, column, gradient programs, and solvents), mass
107 spectrometric method (acquisition mode and calibration method) are presented in the "**Sample**
108 **Information**" sheet in the supporting information (SI) excel spreadsheet. Further, a more detailed
109 description of the samples and methods used are presented in the SI spreadsheet, including information
110 on any previously published datasets.

111 A wide variety of environmental samples were included in this study. The majority of the samples were
112 wastewater (effluent and influent), surface water, and groundwater samples. More than half of the samples
113 (26 out of 48) were wastewater samples (mainly effluent wastewater samples). Wastewater sample data
114 sets were from Switzerland (various locations)¹⁴, Norway, Sweden, Finland, Denmark, Iceland, Spain,
115 Greece, Mexico and Australia. Fifteen of the 48 samples were samples from ecologically important large
116 rivers such as Danube (station JDS57 Bulgarian/Romanian borders)⁷ and Rhine¹⁵, smaller rivers such as

117 Swiss rivers (Furtbach and Doubs)¹⁶, Dutch rivers (Meuse and Vecht) and the Logan river in Australia. Four
118 groundwater samples were included from Spain and the Netherlands. One primary sludge sample from the
119 wastewater treatment plant (WWTP) in Athens (Greece)¹⁷ as well as one seawater sample affected by
120 treated wastewater¹⁸ were also evaluated. Finally, two drinking water samples from Ridderkerk and
121 Lekkerkerk in The Netherlands were included in the study. All the participants were asked to provide only
122 the absolute intensity of the identified features that were blank subtracted in order to avoid the false
123 positive identification.

124 Participating laboratories analyzed their samples using their own routines (i.e. sample preparation and data
125 processing) for all the analytes included in the NormaNEWS suspect list (“**NormaNEWS compounds**” sheet
126 in the SI, on the NORMAN [Suspect Exchange](#) and in the CompTox [Chemistry Dashboard](#)). No specific
127 method (i.e. chromatographic, ion source, and polarity) was recommended to the participants. This was in
128 order to test the applicability of this approach for the data generated via different methods. For these
129 analyses, a wide range of mass analyzers as well as chromatographic conditions was employed by different
130 participants (“**Sample Information**” sheet in the SI). All of the reported results were further examined,
131 through a quality control assessment, to produce harmonized and comparable results (see section ‘Quality
132 control criteria’). Finally, each identified peak was assigned with an appropriate confidence level.¹⁹ These
133 quality assurance steps were deemed necessary for interpretation of the results.

134

135 **NormaNEWS suspect list**

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

158 network.com/?q=node/236) and on the CompTox Chemistry Dashboard
159 (https://comptox.epa.gov/dashboard/chemical_lists/normanews).

160 **Quality control criteria**

161 All participants of NormaNEWS exercise were requested to submit their results together with their raw LC-
162 HRMS chromatograms. Raw chromatograms were converted to mzML using ProteoWizard (msconvert
163 module v.3.0.10827).²⁰ For data acquired in data-independent acquisition mode, different collision energy
164 channels were separated using an in-house script (provided in the SI), while lock mass scans were removed.
165 For data-dependent acquisition mode, MS/MS spectra were exported as text files (named “precursor mass
166 retention time”) and were removed from the mzML files. Treated mzML files were converted to CDF files,
167 which are readable from various data analysis software including Bruker DataAnalysis v.4.3. (Bruker
168 Daltonics, Bremen, Germany), which was used here.

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

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

177 **RESULTS AND DISCUSSION**

178 **Quality control assessment**

179 Quality control was performed to ensure that data were generated from well-calibrated instruments and
180 that the data submitted were reliable. The first and most important step of the procedure was to check
181 that the mass difference between the experimental and theoretical mass did not exceed ± 5 mDa, which
182 was considered the maximum tolerable mass error in the provided complex environmental samples.^{21, 22}
183 This was highly relevant in assessing the confidence level assigned to each identified analyte in the list.

184 The mass accuracy quality control is summarized in the SI “**QC_mass accuracy_ppm/ QC_mass**
185 **accuracy_Da**” sheet and the results presented in Figure 1. According to the submitted datasets, Orbitrap
186 mass analyzers showed better mass accuracy performance (absolute average mass error 0.55 mDa)
187 comparing to other TOF instruments (absolute average mass error 0.91 mDa), based on successfully
188 identified compounds. Mass errors are caused by the complexity of the samples, saturation of the detector
189 (see section challenges and recommendations), and the instrument itself (i.e. the age and hardware). LC-
190 HRMS data obtained using LTQ Orbitrap instruments showed lower mass accuracy (absolute average mass
191 error 1.1 mDa) when compared with the LTQ Orbitrap XL (absolute average mass error 0.52 mDa), which
192 showed lower mass accuracy in comparison with the QExactive. We further investigated the effect of
193 instrumentation used on the observed mass accuracies through a non-parametric statistical test Kruskal-
194 Wallis.²³ A Kruskal-Wallis p value > 0.01 indicated the rejection of null-hypothesis and statistical significance
195 of the observed differences in the measured averaged masses. The method used to calibrate each
196 instrument was also considered. LC-HRMS data obtained using a Bruker QTOF were calibrated by injecting

197 the calibrant substance at the beginning of the chromatogram, while data from Waters QTOF (in both
198 cases) were calibrated by lock-mass every 0.5 or 2 minutes (injecting, recording and recalibrating based on
199 calibrant peaks appearing every 0.5/2 minutes). High mass accuracy is an extremely crucial parameter to
200 achieve high quality results during the suspect analysis. Especially, high accuracy measurements enable a
201 decreased number of false positive detections.

202 The chromatographic stability of the LC separation was also assessed. All participants submitted at least 3
203 datasets for evaluation. Retention time data from the same instrumental set-up (and same partner) were
204 grouped together and the normalized standard deviations (NSD) of the retention times of the detected
205 substances were calculated (retention times of the detected substances in seconds can be found in the SI
206 “**QC_observed_ret.time_Minutes**” sheet). A criterion of the maximum tolerable NSD of 10% was adopted
207 for accepting the detection of a single compound across samples in data coming from the same partner.
208 The average normalized standard deviation of retention times in all samples was < 2% (Figure S1). The
209 largest variability of 8.6 % was observed for analyte valsartan, whereas the lowest variability (<0.1%) was
210 observed for acesulfame in samples from Netherlands, GES-07 in samples from Australia, and GES-09 and
211 GES-06 in samples from Greece. Retention time stability was considered as another extremely important
212 parameter, which has a direct effect on the identification confidence. The low deviation observed in all the
213 submitted datasets indicated the high quality and reliability of the LC separation of the participating
214 laboratories.

215 The third QC criterion related to the presence of qualifier ions (QI) in the MS/MS spectra (SI “**NormaNEWS**
216 **compounds**” sheet). These ions are fragments of the parent ion and are observable at higher collision
217 energy or even at low collision energy as in-source fragments. The criterion was set on the presence of the
218 QIs as either an in-source fragment or at higher collision energy. The identification level of compounds that
219 did not comply with the third QC criterion were regarded as questionable and were marked accordingly.¹⁹
220 As these QIs proved to be a very efficient way of improving the confidence of the suspect hit, Top 3
221 fragments have now been extracted from all mass spectra submitted to [MassBank.EU](#) and also put on the
222 [NORMAN Suspect Exchange \(direct download\)](#) and the [CompTox Chemistry Dashboard Downloads \(direct link\)](#)
223 for community use. The QC stage was used to exclude the features that did not meet the previously
224 set criteria, thus harmonization. Consequently, we have reported only the features that met these
225 mentioned criteria.

226 **Overview of the retrospective screening**

227 PolyEthylene Glycol 09 (PEG-09) was the most frequently detected compound, being present in 41 out of
228 the 48 samples (85%) analyzed. Several bisphenols, transformation products of perfluorooctane sulfonate,
229 and the pharmaceutical omeprazole were not detected in any of the samples analyzed (“**Max. Absolute**
230 **Intensity_counts**” sheet in the SI and Figures 2, XS, X1S, X2S). Series of surfactants, such as [PEGs](#), [C12AEO-](#)
231 [PEGs](#), and [GES](#), resulted in a higher detection frequency for compounds with masses varying between 400
232 and 600 Da compared to both smaller and larger molecules from the same families (Figure S2.A).
233 Schymanski et al and Gago-Ferrero et al. have previously observed a similar trend for these surfactants.¹⁴
234 ²⁴ The observed trend may be explained by the efficient ionization of mid-size molecules compared to
235 other compounds and potentially the fact that they are used as technical mixtures.²⁵ [LAS](#) had an average
236 frequency of detection of around 50%. The largest measured [LAS](#), in terms of mass (i.e. C14-LAS), were
237 detected in only 4 samples out of 48 samples. Based on the estimated retention time for LAS-C14, we

238 interpret that the chromatographic run times used by different partners were not sufficiently long to
239 successfully detect this suspect analyte in the evaluated samples. Only 3 of the 5 suspect fluorinated
240 surfactants were detected with perfluorooctane sulfonate (PFOS) having the highest detection frequency
241 of ~ 35%. For industrial chemicals and pharmaceuticals, venlafaxine was the suspect analyte with the
242 highest frequency of detection (68%), while several bisphenols were not detected in any of the samples.
243 Additionally, we observed a higher occurrence frequency of the suspect analytes in the locations with
244 higher population density such as Spain, Switzerland, and Greece compared to locations such as
245 Scandinavia and Australia with lower population density, Figures 2 and S3. The observed trend was
246 consistent across all the analyzed matrices. However, it should be noted that considering the limited data
247 set for this pilot study, further interpretation of the spatial and temporal distribution of pollutants is not
248 possible. The future implementation of this approach will provide larger datasets for comprehensive spatial
249 and temporal assessment of CEC occurrence across the globe.

250 The presence of a large number of successfully detected surfactants and industrial chemicals in both
251 wastewater influents, effluents, and surface waters suggests the wide spread occurrence of these CECs in
252 the environment across the globe, Figure 2. Although modern wastewater treatment plants are to some
253 extent equipped to remove these pollutants²⁶⁻²⁹, the high production/consumption volumes of these
254 chemicals used in households and industrial applications translates into their release into the environment.
255 The environmental occurrence, fate and behavior of surfactants have been widely investigated, however
256 more reliable environmental data for these pollutants are necessary.³⁰⁻³² Collective exercises such as
257 NormaNEWS are therefore an important step forward towards producing a comprehensive and reliable
258 database on the environmental occurrence of surfactants and/or other chemicals of emerging concern
259 (CEC), which can be used for better understanding of their environmental fate and behavior. Furthermore,
260 this exercise, through the provided QC criteria, metadata template (i.e. SI spreadsheet), provides all
261 necessary information and guidelines for laboratories across the globe for the reliable detection,
262 identification, and reporting of CECs in different environmental compartments.

263 **Challenges and recommendations**

264 For analysts to obtain high-confidence identifications through retrospective suspect screening they face
265 several challenges. Here, recommendations for dealing with difficulties such as broad peaks, data
266 acquisition, and sensitivity are provided in the following.

267 The presence of broad peaks in the chromatograms of complex samples is often caused by the physico-
268 chemical properties of that compound and the selected chromatographic method is unavoidable. For
269 example, the [LAS](#) surfactants that elute at the end of the gradient of a typical reverse phase
270 chromatographic run result in characteristic broad peaks (Figure 3A). Many peak picking algorithms are
271 unable to detect such broad peaks. Therefore, employing peak picking independent approaches^{33,34}, prior
272 knowledge of those analytes, and visualization tools, even though not comprehensive, may be useful in
273 dealing with broad peaks.

274 Data-dependent acquisition is often used in non-target analysis. Certain limitations with data-dependent
275 acquisition may potentially cause false identification of features due to its limitations. This acquisition
276 mode isolates and provides MS/MS spectra of some of the most abundant ions per full scan. Even though
277 this approach is the ideal acquisition mode during identification of peaks with the most abundant ions, this
278 mode is not suitable for retrospective screening, due to the limited number of MS/MS spectra obtained. In

279 case the peak of an environmentally relevant compound is not one of those most abundant ions, the
280 MS/MS spectra of this chemical would not be recorded (Figure 3B). Therefore, confident identification of
281 that peak would not be possible. As a solution, it is highly recommended that samples are injected in data-
282 independent acquisition mode which is the ideal acquisition mode for retrospective screening. In data-
283 independent acquisition, HRMS is recording full scan and MS/MS spectra without prior isolation of any
284 mass. Therefore, all fragments (and fragments of fragments in case of in-source fragments) of all co-eluting
285 compounds are recorded, resulting in complex but information-rich MS/MS spectra that requires adequate
286 data processing tools for confident identification of features. However, to our knowledge this is the most
287 effective acquisition method for the samples that are meant for retrospective analysis. As different
288 compounds have different fragmentation behavior depending on the different collision energies, the use
289 of multiple (e.g. low, medium, high) or ramped collision energies should be considered during acquisition
290 of data for retrospective screening to cover as many compounds as possible. As different instruments have
291 different settings and acquisition speeds, a compromise may need to be found to provide sufficient
292 resolution in the full scan while obtaining as much fragmentation information as possible. Pilot studies such
293 as these and the upload of corresponding suspect lists and fragment information to public resources greatly
294 help exchange experience to find these ideal compromises for future investigations.

295 Another inherent concern about LC-HRMS data is sensitivity. Among other reasons, one possible case for
296 non-detection of pollutants is that current HRMS instruments operated in full scan are sensitive depending
297 on the frequency with which they acquire full scans.³⁵ This means that low abundant or poorly ionized
298 chemicals are not detected in case HRMS instrument records full scans at a high frequency rate. For
299 example, recording full-scans at low frequency (2 Hz) will enable the detection of more compounds in
300 comparison with a higher frequency rate (i.e. 20 Hz). Therefore, the analysts should try to find a
301 compromise between the sampling speed and the sensitivity required for the analyses. For the samples,
302 that are meant to be analyzed via retrospective screening a lower sampling frequency is recommended
303 given that under these conditions a higher sensitivity is achieved.

304 Substances at high concentration levels in extracts and/or having high ionization efficiency can often result
305 in the detector becoming saturated (Figure 3C). In this case, the peak reaches a plateau, which makes peak
306 picking and determination of exact mass and retention time very difficult. For example, surfactants such as
307 [PEGs](#) and [C12AEO-PEGs](#) were affected by detector saturation due to their high concentrations in the
308 evaluated samples. The mentioned uncertainties in the exact mass and retention time are caused by the
309 fact that saturation reduces the mass accuracy of the measurements for certain instruments, which is of
310 extreme importance when performing identification. However, increasing the mass extraction window may
311 solve these issues. On the other hand, such less strict mass accuracy criterion may increase the likelihood
312 of false positive detection.

313 Another open issue in mass spectrometry is related to structural isomers (Figure 3D). Isomers are
314 structurally similar compounds with the same molecular formula (same mass and isotopic profile) and share
315 very similar fragmentation. This happened in the case of the detection of bisphenol S in the surface waters
316 of the Netherlands. Two peaks, with different retention times, with acceptable mass accuracy, isotopic fit
317 and same qualifier ions seem to belong to two different isomers of bisphenol S. In such cases, deeper
318 knowledge of fragmentation behavior and/or retention time prediction could help to identify the peak that
319 belongs to the suspected substance. Ion ratio (ratio of the intensity of a fragment to the intensity of another

320 fragment) can be also considered. However, this information should be carefully examined, because of ion
321 suppression caused by high background signal produced by complex sample's matrix. Classes of substances
322 such as the surfactants mentioned here also contain many structurally related substances that cannot be
323 distinguished easily with mass spectrometry. These are now being grouped as "related substances" in the
324 CompTox Chemistry Dashboard (see hyperlinks for the different surfactant classes throughout this
325 manuscript) as a first step in working towards computational solutions to deal with the extremely complex
326 challenge of chemical substances of Unknown or Variable Composition, Complex Reaction Products and
327 Biological Materials (UVCBs).^{36,37} Finally, all the samples need to be analyzed both in positive and negative
328 mode in order to cover a wider chemical space compared to only single polarity.

329 **The early warning system and its potential**

330 This exercise confirmed the high occurrence frequency of several surfactants (e.g. [PEGs](#) and [C12AEO-PEGs](#)),
331 transformation products of selected drugs (e.g. gabapentin-lactam, metoprolol-acid, carbamazepine-10-
332 hydroxy, omeprazole-4-hydroxy-sulphide, 2-benzothiazole-sulfonic-acid), and industrial chemicals such as
333 3-nitrobenzenesulfonate and bisphenol S. These chemicals are not typically included in target/suspect lists
334 used for surface water monitoring programs. Subsequently, there are limited environmental occurrence
335 data available for these pollutants.³⁸⁻⁴⁰ This clearly demonstrates that an early warning network such as
336 NormaNEWS enables the efficient and reliable detection and identification of novel CECs in different
337 environmental compartments at both a temporal and spatial scale. Consequently, a reasonably large and
338 diverse dataset on the environmental occurrence of novel CECs in different matrices has been generated
339 during this pilot project. Clearly, this study was a proof of concept to test the applicability of such an
340 approach to a diverse global dataset. Further development and larger global coverage is necessary in order
341 to generate a dataset suitable for both environmental interpretation and policy making practices. Such a
342 dataset provides an initial screen that can be used to inform contaminant prioritization exercises leading
343 to further monitoring, fate and effect studies and subsequent risk assessment. Furthermore, given that the
344 data are harmonized across a large number of laboratories and the confidence level of each identification
345 is provided, the inherent reliability of each identification becomes more intuitive to non-experts. The
346 purpose of this network activity would not be to replace ongoing targeted monitoring and screening
347 programs, but to provide a robust and comprehensive complementary collaborative approach for
348 informing the refinement of priority substance lists. This also shows the great potential for screening much
349 larger lists in the future, although the manual verification of the results is still a demanding task. More
350 computationally efficient methods will be needed before this can be expanded to potentially lists of tens
351 of thousands of substances.

352 The NormaNEWS pilot was performed using a very simple approach where all participants manually
353 submitted data on their CECs of interest in order to create a suspect screening list for the collaborative
354 exercise. This enabled researchers to easily obtain additional data on the CECs that they are particularly
355 interested in. Future lists could be generated by a number of different approaches including from open
356 resources, such as massbank.eu. As highlighted recently by Schymanski and Williams,³⁶ open resources will
357 be instrumental in defining the evolution of suspect screening. The community-wide sharing of CECs
358 through the exchange of suspect lists (e.g. the [NORMAN Suspect Exchange](#) and the [Chemistry Dashboard](#)
359 [lists](#)) as well as tentatively and unequivocally identified spectra and sharing the related fragments is
360 therefore key to the success of a global early warning network. Also key will be the willingness of the
361 scientific community to share their HRMS data in an open MS format (e.g. mzML⁴¹, mzXML⁴², and netCDF⁴³).

362 The Global Natural Products Social Molecular Networking (GNPS; <http://gnps.ucsd.edu/>) provides a vision
363 as to how global collaboration and social cooperation can be used to address major scientific challenges in
364 the sharing and community curation of MS data.⁴⁴ Taking inspiration from GNPS, we propose that HRMS
365 data are made available (through a virtual repository and with necessary metadata) in order to facilitate
366 living data along with periodic automated re-analysis of data (i.e. with updates to the suspect list or the
367 addition of new data sets). Ideally, this repository will be easily accessible through a web-application and
368 free of the aforementioned challenges. The environmental and exposure sciences currently lag behind
369 other fields, such as proteomics⁴⁵, metabolomics⁴⁶ and natural product research⁴⁷ in globally collaborating
370 and sharing data through open/social platforms in order to revolutionize the way data are processed to
371 achieve significant outcomes. We acknowledge that not all the data tools are currently in place to make
372 our proposal a reality, however progress is being made in this area^{33, 34, 48, 49}. For example, within the
373 NORMAN Network (<http://www.norman-network.net/>) there is an initiative to develop a digital sample
374 freezing platform. A global emerging contaminant early warning network based on adopting the successful
375 practices of other similar networks will play a pivotal role in identifying chemicals using HRMS that has the
376 potential to possess significant outcomes in protecting human and environmental health.

377 SUPPORTING INFORMATION

378 Text, tables and figures with detailed information on experimental methods, QA/QC procedures
379 and supplemental data (xls, PDF).

380 ACKNOWLEDGEMENTS

381 Support by the NORMAN Network of reference laboratories, research centers and related organizations for
382 the monitoring of emerging environmental substances facilitated this work under its Joint Program of
383 Activities (<http://www.norman-network.net>) and is greatly appreciated. We are especially grateful to
384 Antony J. Williams (US EPA, CompTox Chemistry Dashboard) for integrating the NormaNEWS Suspect List,
385 other NORMAN lists and the fragment data into the Dashboard, as well as for all his efforts in cross-linking
386 the surfactants. JH, HPS and ELS acknowledge the efforts of Philipp Longrée, Heinz Singer and other
387 colleagues at Eawag who contributed to the analysis of the Eawag data. SS and KVT were supported in part
388 by the Research Council of Norway (Project Number 243720). KVT, JFM and SK gratefully acknowledge the
389 financial support of the Queensland Department of Health. ELS was supported in part by the EU FP7 Project
390 SOLUTIONS (Grant Number 603437).

391 REFERENCES

- 392 1. Kortenkamp, A.; Faust, M.; Scholze, M.; Backhaus, T., Low-level exposure to multiple chemicals:
393 reason for human health concerns? *Environ Health Perspect* **2007**, *115 Suppl 1*, 106-114.
- 394 2. Pleil, J. D., Categorizing Biomarkers of the Human Exposome and Developing Metrics for Assessing
395 Environmental Sustainability. *Journal of Toxicology and Environmental Health, Part B* **2012**, *15*, (4), 264-
396 280.
- 397 3. Muir, D. C. G.; Howard, P. H., Are There Other Persistent Organic Pollutants? A Challenge for
398 Environmental Chemists. *Environmental Science & Technology* **2006**, *40*, (23), 7157-7166.

- 399 4. Rager, J. E.; Strynar, M. J.; Liang, S.; McMahan, R. L.; Richard, A. M.; Grulke, C. M.; Wambaugh, J.
400 F.; Isaacs, K. K.; Judson, R.; Williams, A. J.; Sobus, J. R., Linking high resolution mass spectrometry data with
401 exposure and toxicity forecasts to advance high-throughput environmental monitoring. *Environ Int* **2016**,
402 *88*, 269-280.
- 403 5. Andra, S. S.; Austin, C.; Patel, D.; Dolios, G.; Awawda, M.; Arora, M., Trends in the application of
404 high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the
405 environmental chemical space of the human exposome. *Environ Int* **2017**, *100*, 32-61.
- 406 6. Leendert, V.; Van Langenhove, H.; Demeestere, K., Trends in liquid chromatography coupled to
407 high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic
408 environments. *TrAC Trends in Analytical Chemistry* **2015**, *67*, 192-208.
- 409 7. Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.;
410 Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibanez, M.; Portoles, T.; de
411 Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.;
412 Bogialli, S.; Stipanicev, D.; Rostkowski, P.; Hollender, J., Non-target screening with high-resolution mass
413 spectrometry: critical review using a collaborative trial on water analysis. *Anal Bioanal Chem* **2015**, *407*,
414 (21), 6237-55.
- 415 8. Krauss, M.; Singer, H.; Hollender, J., LC-high resolution MS in environmental analysis: from target
416 screening to the identification of unknowns. *Anal Bioanal Chem* **2010**, *397*, (3), 943-51.
- 417 9. Hernandez, F.; Sancho, J. V.; Ibanez, M.; Abad, E.; Portoles, T.; Mattioli, L., Current use of high-
418 resolution mass spectrometry in the environmental sciences. *Anal Bioanal Chem* **2012**, *403*, (5), 1251-64.
- 419 10. Gomez-Ramos, M. M.; Ferrer, C.; Malato, O.; Aguera, A.; Fernandez-Alba, A. R., Liquid
420 chromatography-high-resolution mass spectrometry for pesticide residue analysis in fruit and vegetables:
421 screening and quantitative studies. *J Chromatogr A* **2013**, *1287*, 24-37.
- 422 11. Polgar, L.; Garcia-Reyes, J. F.; Fodor, P.; Gyepes, A.; Dernovics, M.; Abranko, L.; Gilbert-Lopez, B.;
423 Molina-Diaz, A., Retrospective screening of relevant pesticide metabolites in food using liquid
424 chromatography high resolution mass spectrometry and accurate-mass databases of parent molecules
425 and diagnostic fragment ions. *J Chromatogr A* **2012**, *1249*, 83-91.
- 426 12. Chiaia-Hernandez, A. C.; Krauss, M.; Hollender, J., Screening of lake sediments for emerging
427 contaminants by liquid chromatography atmospheric pressure photoionization and electrospray
428 ionization coupled to high resolution mass spectrometry. *Environ Sci Technol* **2013**, *47*, (2), 976-86.
- 429 13. Gomez-Ramos Mdel, M.; Perez-Parada, A.; Garcia-Reyes, J. F.; Fernandez-Alba, A. R.; Aguera, A.,
430 Use of an accurate-mass database for the systematic identification of transformation products of organic
431 contaminants in wastewater effluents. *Journal of chromatography. A* **2011**, *1218*, (44), 8002-12.
- 432 14. Schymanski, E. L.; Singer, H. P.; Longree, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripolles Vidal, C.;
433 Hollender, J., Strategies to characterize polar organic contamination in wastewater: exploring the
434 capability of high resolution mass spectrometry. *Environ Sci Technol* **2014**, *48*, (3), 1811-8.
- 435 15. Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P., Quantitative target and systematic non-target
436 analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry-
437 Identification of unknown sources and compounds. *Water Res* **2015**, *87*, 145-54.
- 438 16. Moschet, C.; Wittmer, I.; Simovic, J.; Junghans, M.; Piazzoli, A.; Singer, H.; Stamm, C.; Leu, C.;
439 Hollender, J., How a complete pesticide screening changes the assessment of surface water quality.
440 *Environ Sci Technol* **2014**, *48*, (10), 5423-32.
- 441 17. Gago-Ferrero, P.; Borova, V.; Dasenaki, M. E.; Tauhomaidis Nu, S., Simultaneous determination of
442 148 pharmaceuticals and illicit drugs in sewage sludge based on ultrasound-assisted extraction and liquid
443 chromatography-tandem mass spectrometry. *Anal Bioanal Chem* **2015**, *407*, (15), 4287-97.
- 444 18. Alygizakis, N. A.; Gago-Ferrero, P.; Borova, V. L.; Pavlidou, A.; Hatzianestis, I.; Thomaidis, N. S.,
445 Occurrence and spatial distribution of 158 pharmaceuticals, drugs of abuse and related metabolites in
446 offshore seawater. *Sci Total Environ* **2016**, *541*, 1097-105.

- 447 19. Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J., Identifying
448 small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol*
449 **2014**, *48*, (4), 2097-8.
- 450 20. Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.;
451 Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.;
452 Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.;
453 Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.;
454 Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss,
455 M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*
456 **2012**, *30*, (10), 918-20.
- 457 21. Zedda, M.; Zwiener, C., Is nontarget screening of emerging contaminants by LC-HRMS successful?
458 A plea for compound libraries and computer tools. *Anal Bioanal Chem* **2012**, *403*, (9), 2493-502.
- 459 22. Kaufmann, A.; Walker, S., Evaluation of the interrelationship between mass resolving power and
460 mass error tolerances for targeted bioanalysis using liquid chromatography coupled to high-resolution
461 mass spectrometry. *Rapid Commun Mass Spectrom* **2013**, *27*, (2), 347-56.
- 462 23. Breslow, N., A generalized Kruskal-Wallis test for comparing K samples subject to unequal
463 patterns of censorship. *Biometrika* **1970**, *57*, (3), 579-594.
- 464 24. Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S.,
465 Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw
466 Wastewater with LC-HRMS/MS. *Environ Sci Technol* **2015**, *49*, (20), 12333-41.
- 467 25. Mazzoni, M.; Rusconi, M.; Valsecchi, S.; Martins, C. P.; Polesello, S., An on-line solid phase
468 extraction-liquid chromatography-tandem mass spectrometry method for the determination of
469 perfluoroalkyl acids in drinking and surface waters. *J Anal Methods Chem* **2015**, *2015*, 942016.
- 470 26. Prats, D.; Ruiz, F.; Vazquez, B.; M., R.-P., Removal of anionic and nonionic surfactants in a
471 wastewater treatment plant with anaerobic digestion. A comparative study. *Water Res* **1997**, *31*, (8),
472 1925-1930.
- 473 27. Aboulhassan, M. A.; Souabi, S.; Yaacoubi, A.; Baudu, M., Removal of surfactant from industrial
474 wastewaters by coagulation flocculation process. *Int J Environ Sci Tech* **2006**, *3*, (4), 327-332.
- 475 28. Gonzalez, S.; Petrovic, M.; Barcelo, D., Removal of a broad range of surfactants from municipal
476 wastewater--comparison between membrane bioreactor and conventional activated sludge treatment.
477 *Chemosphere* **2007**, *67*, (2), 335-43.
- 478 29. Luo, Y.; Guo, W.; Ngo, H. H.; Nghiem, L. D.; Hai, F. I.; Zhang, J.; Liang, S.; Wang, X. C., A review on
479 the occurrence of micropollutants in the aquatic environment and their fate and removal during
480 wastewater treatment. *Sci Total Environ* **2014**, *473-474*, 619-41.
- 481 30. Jackson, M.; Eadsforth, C.; Schowanek, D.; Delfosse, T.; Riddle, A.; Budgen, N., Comprehensive
482 review of several surfactants in marine environments: Fate and ecotoxicity. *Environ Toxicol Chem* **2016**,
483 *35*, (5), 1077-86.
- 484 31. Jardak, K.; Drogui, P.; Daghrir, R., Surfactants in aquatic and terrestrial environment: occurrence,
485 behavior, and treatment processes. *Environ Sci Pollut Res Int* **2016**, *23*, (4), 3195-216.
- 486 32. Ying, G.-G., Fate, behavior and effects of surfactants and their degradation products in the
487 environment. *Environment International* **2006**, *32*, (3), 417-431.
- 488 33. Samanipour, S.; Langford, K.; Reid, M. J.; Thomas, K. V., A two stage algorithm for target and
489 suspect analysis of produced water via gas chromatography coupled with high resolution time of flight
490 mass spectrometry. *J Chromatogr A* **2016**, *1463*, 153-61.
- 491 34. Samanipour, S.; Baz-Lomba, J. A.; Alygizakis, N. A.; Reid, M. J.; Thomaidis, N. S.; Thomas, K. V., Two
492 stage algorithm vs commonly used approaches for the suspect screening of complex environmental
493 samples analyzed via liquid chromatography high resolution time of flight mass spectroscopy: A test study.
494 *J Chromatogr A* **2017**, *1501*, 68-78.

- 495 35. Acena, J.; Stampachiachchiere, S.; Perez, S.; Barcelo, D., Advances in liquid chromatography-high-
496 resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal Bioanal Chem*
497 **2015**, *407*, (21), 6289-99.
- 498 36. Schymanski, E. L.; Williams, A. J., Open Science for Identifying "Known Unknown" Chemicals.
499 *Environ Sci Technol* **2017**, *51*, (10), 5357-5359.
- 500 37. Williams A.; Grulke, C. M.; McEachran A.; Richard, A.; Jolley R; Dunne J; Edmiston E; J, E. Comptox
501 Chemistry Dashboard: Web-based data integration hub for environmental chemistry and toxicology data.
502 [https://www.slideshare.net/AntonyWilliams?utm_campaign=profiletracking&utm_medium=sssite&utm](https://www.slideshare.net/AntonyWilliams?utm_campaign=profiletracking&utm_medium=sssite&utm_source=ssslideview)
503 [source=ssslideview](https://www.slideshare.net/AntonyWilliams?utm_campaign=profiletracking&utm_medium=sssite&utm_source=ssslideview)
- 504 38. Beretsou, V. G.; Psoma, A. K.; Gago-Ferrero, P.; Aalizadeh, R.; Fenner, K.; Thomaidis, N. S.,
505 Identification of biotransformation products of citalopram formed in activated sludge. *Water Res* **2016**,
506 *103*, 205-14.
- 507 39. Nika, M. C.; Bletsou, A. A.; Koumaki, E.; Noutsopoulos, C.; Mamais, D.; Stasinakis, A. S.; Thomaidis,
508 N. S., Chlorination of benzothiazoles and benzotriazoles and transformation products identification by LC-
509 HR-MS/MS. *J Hazard Mater* **2017**, *323*, (Pt A), 400-413.
- 510 40. Christophoridis, C.; Nika, M. C.; Aalizadeh, R.; Thomaidis, N. S., Ozonation of ranitidine: Effect of
511 experimental parameters and identification of transformation products. *Sci Total Environ* **2016**, *557-558*,
512 170-82.
- 513 41. Martens, L.; Chambers, M. C.; Sturm, M.; Kessner, D.; Levander, D.; Shofstahl, J.; Tang, W. H.;
514 Römpf, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.;
515 Souda, P.; Hermjakob, H.; Binz, P.-A.; Deutsch, E. W., mzML - a Community Standard for Mass Spectrometry
516 Data. *Mol Cell Proteomics* **2011**, *10*, (1).
- 517 42. Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson,
518 E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp,
519 E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.;
520 Aebersold, R., A common open representation of mass spectrometry data and its application to
521 proteomics research. *Nat Biotechnol* **2004**, *22*, (11), 1459-66.
- 522 43. Erickson, B., ANDI MS standard finalized. *Anal Chem* **2000**, *72*, (3), 103 A-103 A.
- 523 44. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous,
524 J.; Kaponov, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.;
525 Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn,
526 R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D. J.; Gavilan, R. G.; Kleigrew, K.; Northen, T.; Dutton, R. J.; Parrot,
527 D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean,
528 J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.; Keyzers, R. A.;
529 Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; P, C. A. B.; Torres-
530 Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand,
531 E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.;
532 Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.;
533 Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins,
534 S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M.
535 C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.;
536 Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer,
537 D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.;
538 Palsson, B. O.; Pogliano, K.; Lington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.;
539 Dorrestein, P. C.; Bandeira, N., Sharing and community curation of mass spectrometry data with Global
540 Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34*, (8), 828-837.

- 541 45. Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-
542 Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O., OpenMS - an open-source software framework for mass
543 spectrometry. *BMC Bioinformatics* **2008**, *9*, 163.
- 544 46. Uppal, K.; Walker, D. I.; Liu, K.; Shuzhao, L.; G., Y.-M.; P., J. D., Computational Metabolomics: A
545 Framework for the Million Metabolome. *Chem Res Toxicol* **2016**, *29*, (12), 1956-1975.
- 546 47. Allard, P. M.; Genta-Jouve, G.; Wolfender, J. L., Deep metabolome annotation in natural products
547 research: towards a virtuous cycle in metabolite identification. *Curr Opin Chem Biol* **2017**, *36*, 40-49.
- 548 48. Samanipour, S.; Reid, M. J.; Thomas, K. V., Statistical Variable Selection: An Alternative
549 Prioritization Strategy during the Nontarget Analysis of LC-HR-MS Data. *Anal Chem* **2017**, *89*, (10), 5585-
550 5591.
- 551 49. Samanipour, S.; Reid, M.; Baek, K.; Thomas, K. V., Combining a deconvolution and a universal
552 library search algorithm for the non-target analysis of data independent LC-HRMS spectra. *Environ Sci*
553 *Technol* **2018**, In Press.
- 554

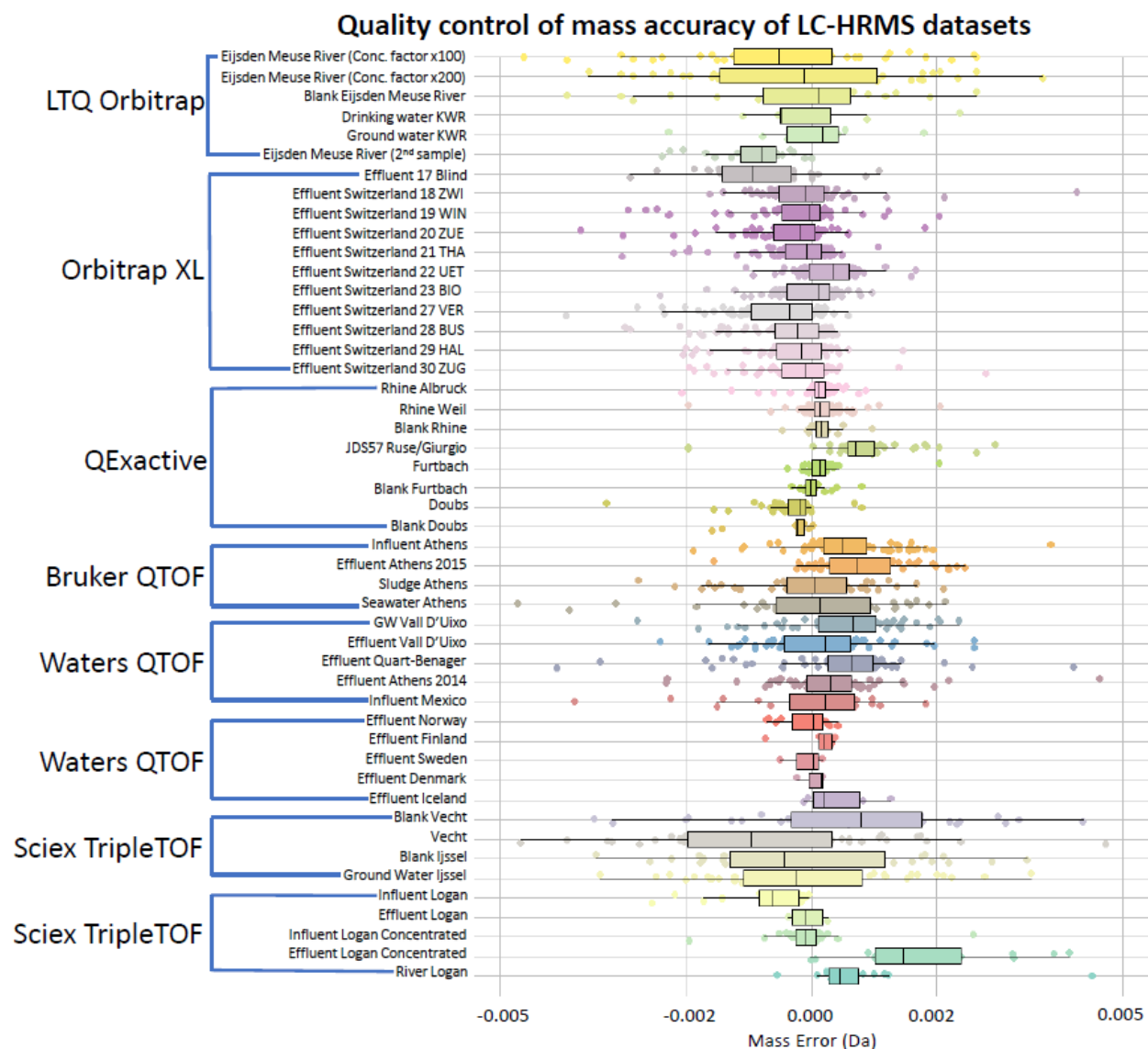
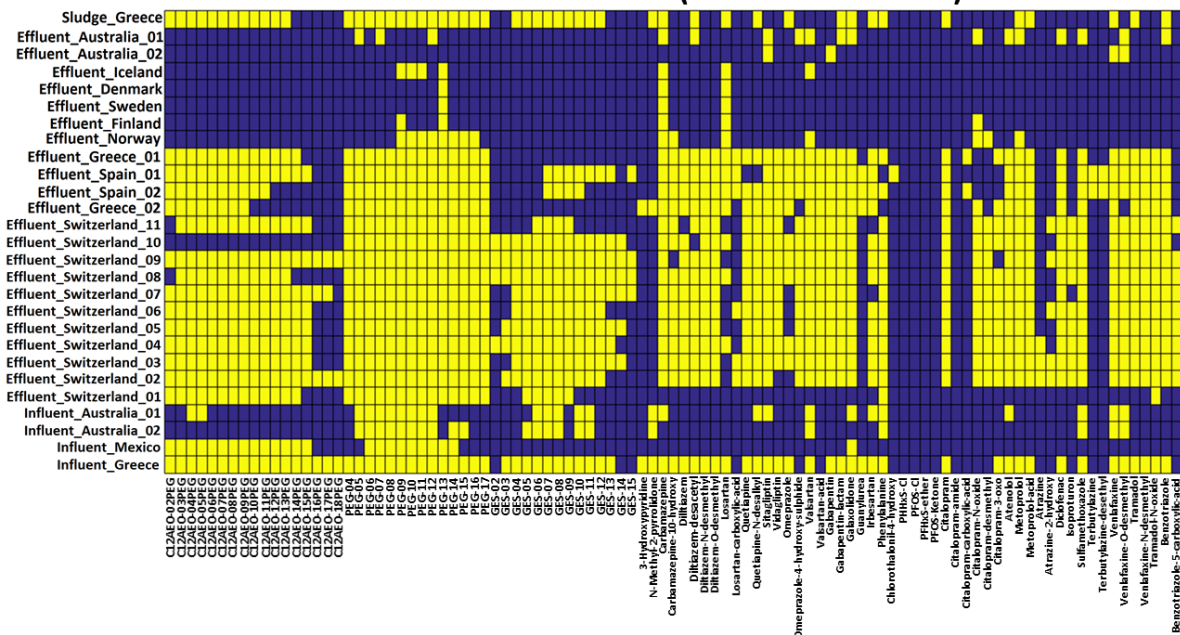


Figure 1. Quality control of mass accuracy of the submitted datasets based on the identified compounds. Type of mass analyzer, calibration type of the mass analyzer as well as other factors (age of equipment, scan sampling rate of the detector) affect the performance and the quality of the results.

Wastewater matrices (Positive Ionization)



Wastewater matrices (Negative Ionization)

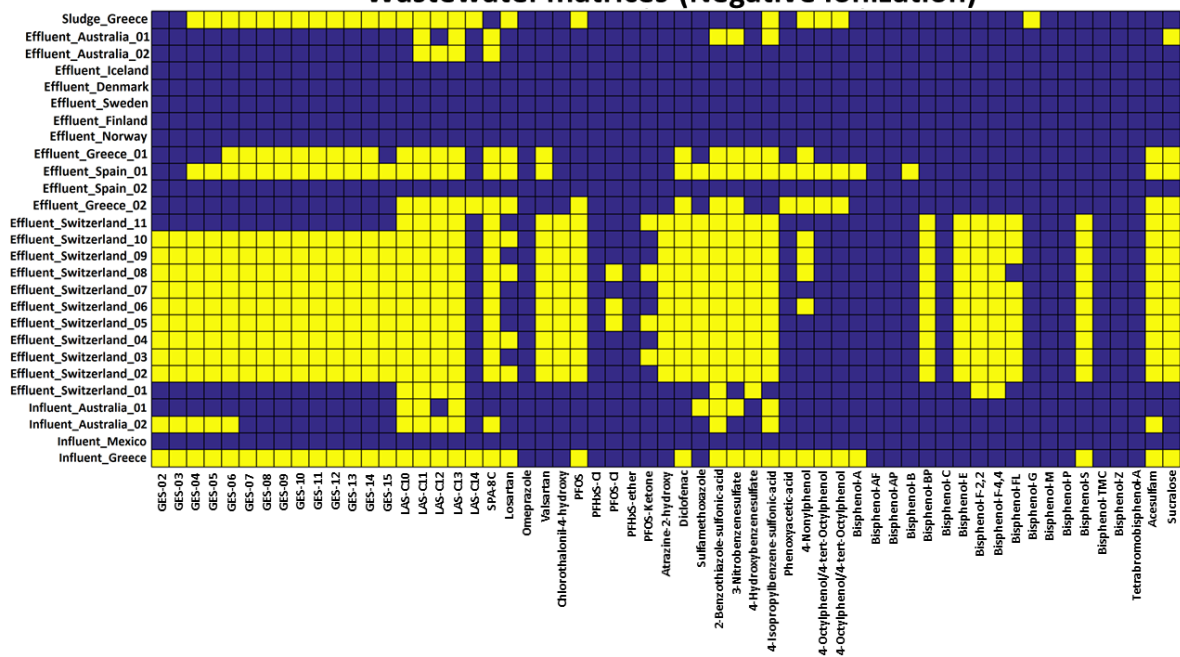


Figure 2. Heat map showing the occurrence of the selected substances in the retrospectively screened samples (primary sludge from WWTP of Athens, Greece, effluent wastewater samples from Australia, Iceland, Spain, Denmark, Sweden, Finland, Norway, Greece and Switzerland) and influent wastewater samples (Australia, Mexico, Greece) for positive and negative ionization. Successfully identified compounds are marked in yellow.

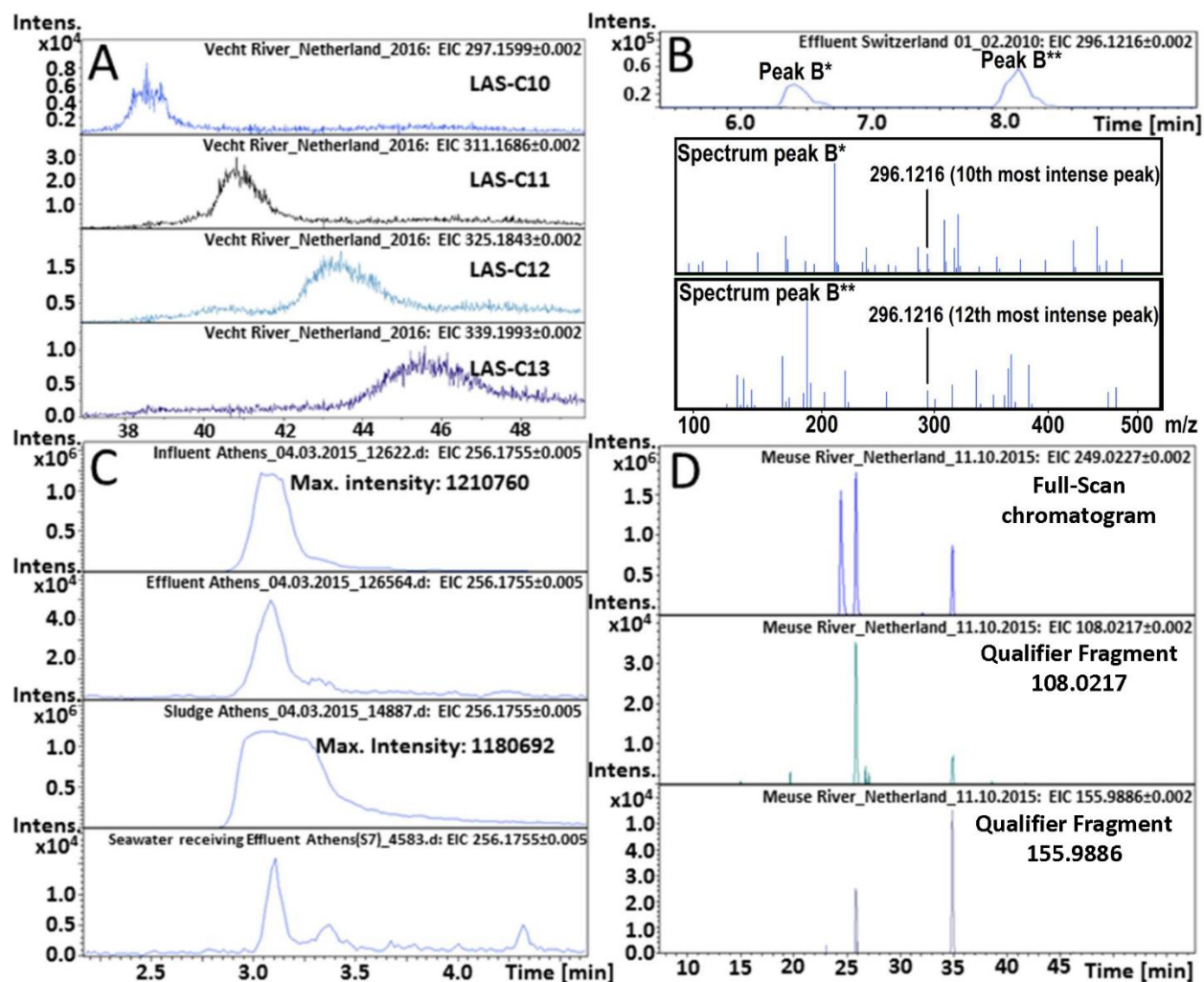


Figure 3. Challenges faced during evaluation of the results; A. Broad peaks of Linear alkylbenzene sulphonate (LAS) surfactants makes peak-picking challenging, B. Missing fragmentation information (MS/MS) of compound of interest decreases identification confidence, because data-dependent acquisition 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.