

Accepted Manuscript

This is an Accepted Manuscript of the following article:

Saer Samanipour, Jose A. Baz-Lomba, Nikiforos A. Alygizakis, Malcolm J. Reid, Nikolaos S. Thomaidis, Kevin 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. *Journal of Chromatography A*. Volume 1501, 2017, Pages 68-78, ISSN 0021-9673.

The article has been published in final form by Elsevier at
<https://doi.org/10.1016/j.chroma.2017.04.040>

© 2017. This manuscript version is made available under the
CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

It is recommended to use the published version for citation.

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

Saer Samanipour^{a,*}, Jose A. Baz-Lomba^a, Nikiforos A. Alygizakis^b, Malcolm J. Reid^a, Nikolaos S. Thomaidis^b, Kevin V. Thomas^{a,c}

^aNorwegian Institute for Water Research (NIVA), 0349 Oslo, Norway

^bLaboratory of Analytical Chemistry, Department of Chemistry, University of Athens, Panepistimiopolis Zographou, 15771 Athens, Greece

^cQueensland Alliance for Environmental Health Science (QAEHS), University of Queensland, 39 Kessels Road, Coopers Plains QLD 4108, Australia

Abstract

LC-HR-QTOF-MS recently has become a commonly used approach for the analysis of complex samples. However, identification of small organic molecules in complex samples with the highest level of confidence is a challenging task. Here we report on the implementation of a two stage algorithm for LC-HR-QTOF-MS datasets. We compared the performances of the two stage algorithm, implemented via NIVA_MZ_AnalyzerTM, with two commonly used approaches (i.e. feature detection and XIC peak picking, implemented via UNIFI by Waters and TASQ by Bruker, respectively) for the suspect analysis of four influent wastewater samples. We first evaluated the cross platform compatibility of LC-HR-QTOF-MS datasets generated

*Saer Samanipour

Email address: saer.samanipour@niva.no (Saer Samanipour)

¹NIVA, Gaustadalléen 21, 0349 Oslo, Norway

Tel: +47 98222087

via instruments from two different manufacturers (i.e. Waters and Bruker). Our data showed that with an appropriate spectral weighting function the spectra recorded by the two tested instruments are comparable for our analytes. As a consequence, we were able to perform full spectral comparison between the data generated via the two studied instruments. Four extracts of wastewater influent were analyzed for 89 analytes, thus 356 detection cases. The analytes were divided into 158 detection cases of artificial suspect analytes (i.e. verified by target analysis) and 198 true suspects. The two stage algorithm resulted in a zero rate of false positive detection, based on the artificial suspect analytes while producing a rate of false negative detection of 0.12. For the conventional approaches, the rates of false positive detection varied between 0.06 for UNIFI and 0.15 for TASQ. The rates of false negative detection for these methods ranged between 0.07 for TASQ and 0.09 for UNIFI. The effect of background signal complexity on the two stage algorithm was evaluated through the generation of a synthetic signal. We further discuss the boundaries of applicability of the two stage algorithm. The importance of background knowledge and experience in evaluating the reliability of results during the suspect screening was evaluated.

Keywords:

Two stage algorithm, LC-HR-QTOF-MS, Target and suspect screening, False detection, Confident identification

1. Introduction

The suspect and non-target analysis of complex environmental samples for small organic molecules is a challenging task [1, 2]. For the samples that are selected *a priori* for suspect and non-target analysis, the analyst performs a generic sample

5 preparation and uses a wide range instrumental analysis [2–6]. A list of suspects is
6 then compiled including some chemical specific information. This information is used
7 for the confident identification of suspect analytes while processing those samples [1–
8 6]. For non-target analysis, the analyst after performing generic sample preparation
9 and instrumental analysis investigates the data without having any prior information
10 about the compounds of interest [2]. Recent advancements in high resolution mass
11 spectrometers coupled to liquid and/or gas chromatography (LC-HR-MS and/or GC-
12 HR-MS) has enabled the analysts to perform these type of analyses faster and with
13 high levels of confidence in the identification, having access to the exact mass in-
14 formation [1, 7]. However, due to the large amount of data produced by these new
15 technologies a confident identification during the suspect and non-target analysis of
16 complex samples remains a difficult and time consuming task.

17

18 Several approaches are available for the suspect analysis of complex samples an-
19 alyzed via LC-HR-MS. However, the majority of these data processing algorithms
20 employ two different strategies in suspect analysis when dealing with the data pro-
21 duced via LC-HR-MS [1, 8]. The first strategy is based on peak picking in the
22 extracted ion chromatograms (XICs) of the molecular ion and the relevant qualifier
23 ions in the sample chromatogram. This is performed by first extracting the XICs
24 and then performing peak picking on those XICs. That way the retention time of
25 a peak is first defined and then matched with other fragment XICs (i.e. qualifiers).
26 In case of a positive match between the XIC retention time of the parent ion and
27 the fragments, those fragments are included in the spectra of that peak. The parent

28 ion and the included fragments, based on their retention time match, are used as
29 diagnostic tools during the identification. Confidence in the identification then is
30 improved by adding another orthogonal parameter to the parent ion and the frag-
31 ments match. These parameters include either the XIC peak shape (employed by
32 ChromaLynx, Waters) or isotopic fit (utilized by TASQ, Bruker). In other words,
33 the algorithms using the XIC strategy, employ XIC retention time match and either
34 isotopic fit or XIC peak shape as two orthogonal parameters in order to identify a
35 suspect peak in the sample. The second data processing strategy consists of first
36 performing peak picking or feature detection in the MS^1 (or first function having a
37 collision energy varying from 4 eV to 6 eV) independently from the target or suspect
38 list (e.g. UNIFI by Waters, MZmine [9], XCMS [10], and OpenMS [11]). Afterwards
39 these algorithms group all the ions related to a specific unique feature. This opera-
40 tion includes grouping the fragments, isotopes, and adducts. Once the unique feature
41 detection is completed, these algorithms use the information provided by the ana-
42 lyst, such as the exact masses of molecular ions and potential fragments to assign an
43 identity to a certain unique component. In the last stage, the analyst needs to create
44 filters based on isotopic fit, theoretical fragmentation pattern, mass accuracy of the
45 parent/fragment ions, and/or retention time match of the standard and the sample
46 peaks to prioritize amongst the potential candidates. The list of potential candidates
47 then are processed using carefully designed filters in order to remove false positives,
48 thus increasing the level of confidence in the final identification. This post processing
49 step is necessary, independently from the detection method used (i.e. peak picking
50 or XICs).

51

52 The commonly used data processing approaches for suspect and non-target anal-
53 ysis are not always as effective as needed, particularly for highly complex samples.
54 One of the frequently observed issues with these algorithms is peak picking or fea-
55 ture detection (as reviewed by Zhang et al.[12] and [13]). Also it has been shown
56 that even parameter optimization [14] may not result in a perfect outcome for all
57 the peaks, due to the sample complexity [14, 15]. Therefore, considering that the
58 majority of algorithms have peak picking as an essential step during data process-
59 ing, they may be affected by the probability of failure during peak picking due to
60 the high levels of noise and broad peaks. Independently from peak picking, another
61 commonly observed issue is caused by mass calibration [16]. Improper mass calibra-
62 tion can create difficulties during the isotopic fit and/or the grouping process. These
63 potential artifacts caused by the complexity of the samples may translate into the
64 erroneous identification of a suspect peak. Overall, these algorithms require careful
65 tuning and post processing in order to produce identifications with a high levels of
66 confidence.

67

68 In our recent study, we developed an alternative algorithm (i.e. two stage algo-
69 rithm, for detail information please see section 2.4.1) for target and suspect screening
70 in complex samples analyzed via high resolution time of flight mass spectrometer
71 (HR-TOF-MS) [17]. The two stage algorithm does not perform peak picking or peak
72 deconvolution in the sample chromatogram, which reduces its probability of failure
73 caused by the peak picking algorithm. This algorithm takes advantage of a full spec-

74 tral comparison between the spectra of the standard stored in a local or a public
75 spectral library and the sample. The two stage algorithm calculates the similarity
76 between the standard spectrum and sample spectrum at each scan number of the
77 sample chromatogram. If a suspect analyte is present in the sample then a peak of
78 similarity is observed in the sample. The two stage algorithm was modified in order
79 to process the data produced via LC-HR-QTOF-MS. This algorithm is shown to be
80 effective in the target and suspect analysis of complex samples with the highest levels
81 of confidence and minimum post processing efforts.

82

83 Herein we report the comparison between three different data processing strate-
84 gies for the suspect analysis of 89 analytes, including pharmaceuticals and illicit
85 drugs, in four wastewater extracts (i.e. 4×89 analytes, 356 detection cases).
86 The investigated strategies comprised of: the two stage algorithm implemented via
87 NIVA_MZ_AnalyzerTM, UNIFI by Waters, and TASQ by Bruker. One hundred and
88 fifty eight detection cases (i.e. 2×33 Oslo analytes + 7 common analytes + 2×32
89 Athens analytes + 7 common analytes) out of 356 total cases were artificial suspects
90 whereas the remaining 198 detection cases were analyzed only in suspect mode. In
91 other words, the presence and/or absence of an artificial suspect analyte in the sam-
92 ples was confirmed through target analysis. In this study, UNIFI was considered a
93 representative of the methods, that perform peak picking on the whole m/z domain
94 whereas TASQ was used as an example for peak picking on XICs. On the other hand,
95 NIVA_MZ_AnalyzerTM, which does not rely on peak picking, was considered as an
96 alternative approach. We used influent wastewater extracts as our testing matrix

97 due to their complexity, high level of noise, and background signal. In this study we
98 evaluated the problems that analysts face during this type of analysis and the way
99 that each algorithm tries to overcome these issues. The cross platform compatibility
100 of the data produced via LC-HR-QTOF-MS as well as the applicability boundaries
101 of the two stage algorithm were also discussed.

102 **2. Experimental**

103 *2.1. Chemicals and sample collection*

104 For our assessment we used a suite of 89 analytes consisting of pharmaceuti-
105 cals (including antibiotics) and illicit drugs. These chemicals were selected based on
106 their importance and high frequency of detection in the wastewater samples from
107 both Oslo [18–20] and Athens [5, 21]. These chemicals were divided in Oslo analytes
108 (33 out of 89), the Athens analytes (32 out of 89), the library analytes (17 out of
109 89), and the common analytes (7 out of 89). A complete list of the suspect/target
110 analytes is included in the Supporting Information, section S1.

111

112 We examined the performances of the three selected approaches on four extracts
113 of wastewater influent. Two out of the four samples were collected from the VEAS
114 treatment plant in Oslo. Throughout this article we will refer to these samples as
115 "Oslo samples". The other two samples of influent were from Athens wastewater
116 treatment plants (hereafter referred to as "Athens samples"). These four samples
117 were 1L flow proportional 24 h composite samples of the two studied wastewater
118 treatment plants. These samples were part of a large sampling campaign conducted

119 during the spring of 2015, where we collected one sample every day. We randomly
120 selected four samples out of the pool of samples for evaluation of data processing
121 approaches. These four samples appeared to be representative of the pool of sam-
122 ples collected during the sampling campaign, based on their chromatogram similarity
123 (TIC) to the other samples collected during the sampling campaign.

124

125 The Oslo plus common analytes, 40 compounds, and the Athens plus common
126 analytes, 39 chemicals, were analyzed in both target and suspect analysis mode in
127 Oslo samples and Athens samples, respectively. The Oslo analytes plus the library
128 analytes were considered as true suspects in Athens samples whereas the Athens
129 analytes and the library analytes were analyzed as pure suspects in Oslo samples.
130 This experimental design enabled us to fully evaluate the capabilities and limitations
131 of all three approaches for both suspect and retrospective analysis.

132 *2.2. Sample preparation*

133 The Oslo samples were extracted using Oasis HLB sorbent [18] whereas the
134 Athens samples were extracted using a modified version [5] of the method devel-
135 oped by Kern et al. 2009 [22].

136

137 We considered these samples adequate, based on their complexity (as reviewed
138 by Krauss et al. [23]), high level of background signal, large number of unique
139 features, and finally the diversity of the compounds present in this type of sample
140 [5, 18, 21, 24]. These characteristics of the samples enabled us to test the capabilities
141 and also the limitations of the tested approaches.

142 *2.3. Instrumental setups*

143 For our analysis we employed two different instrumental setups. Xevo G2-S Q-
144 TOF-MS (Waters, USA) was used for Oslo samples while Maxis Impact Q-TOF-MS
145 (Bruker, Germany) was employed for Athens samples. Each instrumental setup
146 performed the analysis under different data acquisition conditions. The diversity
147 in the instruments and the data acquisition settings introduced an extra level of
148 complexity to our dataset, which was desired in order to better characterize the limits
149 in the capabilities of the investigated strategies. Detailed information regarding
150 the chromatography and HR-MS conditions is provided in Supporting Information
151 section 2.3.

152 *2.4. Data processing algorithms*

153 We evaluated three different approaches used for suspect analysis of complex
154 samples. Among the alternative methods, the two stage algorithm implemented
155 via NIVA_MZ_AnalyzerTM was selected. The two stage algorithm was previously
156 developed for target and suspect analysis of the GC-HR-TOF-MS data [17]. The
157 second tested algorithm was UNIFI by Waters (for more information see section
158 2.4.2). This algorithm was considered a representative method for software that
159 employ feature detection or peak peaking and/or modeling as the first step of analysis
160 (e.g. MZmine [9] and XCMS [10]). The third approach was implemented via TASQ
161 by Bruker. This algorithm utilizes the retention time match of the XICs and the
162 isotopic ratios as a means for identification, similar to software packages such as
163 MassHunter by Agilent and ChromaLynx by Waters. These three software programs
164 appeared to cover the two approaches commonly used and an alternative one for

165 the suspect and target analysis of complex samples analyzed via LC-HR-QTOF-MS.
166 Moreover, this is one of the rare studies where the performance of an open access
167 software is compared to the vendor software packages.

168 2.4.1. *NIVA_MZ_AnalyzerTM*

169 The NIVA_MZ_AnalyzerTM is a Graphical User Interface (GUI) via matlab [25]
170 for implementation of the two stage algorithm [17] applicable to both GC-HR-TOF-
171 MS and LC-HR-QTOF-MS data. This algorithm is also independent from mass
172 spectrometer manufacturers. The two stage algorithm produces a spectra library of
173 the standards in the first stage and then calculates the similarity score between each
174 scan of the sample chromatogram and the standard spectrum. More information
175 regarding the suspect analysis procedure is provided in SI, section S4.1.

176

177 During the first stage (i.e. Unique Ion Extractor) the MS¹ and MS² spectra for
178 each standard peak are processed separately. These two spectra are then recorded
179 and stored in the standard library. Therefore, for each entry in the standard library
180 there are two spectra for the low and high collision energies. During the second stage,
181 the MS¹ and MS² spectra are treated as two complementary spectra. Therefore,
182 the similarity matrix between the library component and the sample spectra are
183 calculated separately for the MS¹ and MS². The similarity matrix is the dot product
184 of the weighted and normalized reference spectra and each scan of the sample [17].
185 Thus if an analyte is present in the sample, higher levels of similarity scores are
186 observed at the same location for both the MS¹ and MS² spectra when compared
187 to the background similarity values. Furthermore, five XICs (including the parent

188 ion and 4 qualifiers) are extracted in order to increase the level of confidence in
189 the identification. The two stage algorithm uses a six point criterion for positive
190 and negative identifications (for more information please see section 2.6 and S4).
191 This algorithm showed to be effective for the suspect screening of GC-HR-TOF-MS
192 data [17] however, this is the first implementation of the two stage algorithm for
193 LC-HR-QTOF-MS data.

194 *2.4.2. UNIFI*

195 UNIFI is a commercially available software from Waters. This software first per-
196 forms the peak picking/feature detection in the whole chromatogram employing the
197 MS¹ signal, without taking into account the analytes of interest. UNIFI, utilizes
198 the continuous wavelet transformations (CWTs) developed by Du et al., 2006 [26].
199 This algorithm (i.e. CWTs) performs the peak detection, peak modeling, and also
200 the noise reduction simultaneously. Further information about the CWTs can be
201 found elsewhere [15]. The next step is the grouping of ions in both the MS¹ and
202 MS² spectra, which belong to a the same compound using the isotopic ratios, 2D
203 feature shape (i.e. both in time and m/z domains), and the feature retention time.
204 This process results in the final peak and/or unique feature list and the deconvoluted
205 spectra of each individual chromatographic peak in the list. During the suspect anal-
206 ysis, UNIFI employs the information such as the exact mass of: the molecular ion,
207 fragments, and the potential adducts for the identification. Finally, UNIFI employs
208 user defined filters to priorities the features that are more likely to be positive de-
209 tections. Work-flows similar to this approach have been widely utilized for suspect
210 and non-target analysis of complex samples as reviewed elsewhere [1, 8, 27].

211

212 *2.4.3. Target Analysis for Screening and Quantitation (TASQTM)*

213 TASQ, is a commercially available software package provided by Bruker for the
214 target and suspect analysis of complex samples [28]. The algorithm implemented in
215 this software package uses the information provided by the analyst (i.e. the exact
216 mass of molecular ion and the fragments or the molecular formula of fragments
217 (which contains the information of their exact mass) in order to extract the XIC
218 of the relevant fragments for the analyzed suspect. In this case this method first
219 performs the peak picking in the XIC of the exact mass provided by the analyst. The
220 theoretical isotopic pattern of the molecular ion in the MS¹ is checked as the second
221 step in the identification. Once the isotopic pattern of the molecular ion was fit, then
222 the algorithm uses the XICs of the fragments provided by the user for increasing
223 the confidence level in the identification. These XICs are grouped together based
224 on the match in their retention times. During the final step, a score is generated
225 for each peak based on the isotopic fit, mass accuracy, and XIC retention times.
226 This methodology has been successfully used for the target and suspect analysis of
227 environmental samples, including wastewater extracts [5].

228 *2.5. Target and suspect analysis*

229 The chromatograms of all four samples were analyzed both in target and suspect
230 analysis mode (see section S3 for more information regarding the data preprocessing).
231 We used NIVA_MZ_AnalyzerTM only in suspect analysis mode for all four samples
232 (i.e. Oslo samples and Athens samples). This implied that we did not use the reten-

233 tion time information for any of 89 analytes investigated in this study for analysis
234 with NIVA_MZ_AnalyzerTM. As a validation tool, the Oslo samples were analyzed
235 via UNIFI both in target and suspect mode for the Oslo analytes (i.e. 33 analytes)
236 plus the common analytes (i.e. 7 analytes). The same samples were analyzed only in
237 suspect mode for the Athens and library analytes. In other words the Oslo samples
238 were processed twice. First in suspect mode with all 89 analytes and a second time
239 in target analysis mode only with the Oslo analytes plus the common analytes. A
240 similar approach was adopted for the Athens samples where we first analyzed those
241 samples in suspect mode for all 89 analytes and in the second round we analyzed
242 the same samples in target mode including only the Athens analytes plus common
243 analytes. This approach enabled the cross validation of the results of our suspect
244 analysis for all three software.

245 *2.6. The identification criteria during the target and suspect analysis*

246 During both target and suspect analysis we employed conservative identification
247 criteria, which reduced the odds of false positive detection. For the NIVA_MZ_AnalyzerTM
248 a suspect analyte was considered present in the sample if it met at least 6 out of
249 7 criteria. These criteria consisted of peaks in the similarity matrix for both the
250 MS¹ and MS² spectra; the peak of the exact mass; the peak for at least 2 out 4
251 XICs; and finally the retention time match for all 5 to 7 peaks. These criteria were
252 established based on our preliminary assessment of the two stage algorithm for the
253 LC-HR-QTOF-MS data.

254

255 UNIFI used a six point criteria for positive and negative detections during the

256 suspect analysis whereas it used a seven point criteria for target analysis. These
257 criteria included the isotopic fit, mass accuracy of ± 2 mDa for all the relevant ions,
258 detection of at least 2 out of 4 fragments, and the XIC retention time match for those
259 ions in the sample chromatogram. These XIC retention times were independent from
260 the analyte retention time and only were used for the spectral deconvolution. During
261 the target analysis for the Oslo analytes, the analyte retention times were compared
262 to the retention times of the potential features in the sample. A retention window
263 of 0.25 min was employed for the target screening, section S4 of Supporting Infor-
264 mation. These settings were previously optimized for target and suspect analysis of
265 wastewater samples [18].

266

267 The positive detection criteria when using TASQ consisted of: a mass accuracy
268 of ± 2 mDa for both the exact mass and the fragments; a good isotopic fit, which
269 was represented with the "mSigma Tolerance" parameter with a maximum of 100;
270 a positive match for at least 2 out of 4 fragments; and fragment retention match, in
271 order to include that fragment in the spectra of the sample peak. Also for TASQ,
272 the retention time of the standards were compared to the retention times of the
273 potential peaks, having a retention tolerance of 0.2 min, during the target analysis.
274 This implied that for an analyte to have its presence confirmed in the sample, dur-
275 ing the suspect analysis, it must have at least 6 out of 7 criteria fulfilled. For the
276 target analysis TASQ matched the standard retention time to the sample peak re-
277 tention time using a retention tolerance of 0.2 min. More information regarding the
278 parameter settings of TASQ are provided in SI, section S4.3. A recent study of the

279 wastewater influent samples from Athens treatment plant showed the applicability
280 of these parameters for our type of analysis [5].

281

282 The fragments used in both TASQ and UNIFI were extracted from the mzCloud
283 database [29]. For analytes where more than one source was available, we gave pri-
284 ority to the average spectra recorded by Eawag. The mzCloud was selected as the
285 reference database for the qualifier ions. This selection was based on the quality of
286 the spectra stored in this database, the fact that this database was an independent
287 database, and the ease of use [30]. We used only the annotated fragments with rel-
288 atively high intensities in the average spectra of each analyte. A text file with the
289 compound names, smiles, exact mass of the parent ion, and the exact mass of four
290 fragments compiled and then imported into both UNIFI and TASQ. Therefore, the
291 absolute intensity of the ions in the reference spectra (i.e. the mzCloud entry) was
292 not considered during the identification via both UNIFI and TASQ.

293

294 For the NIVA_MZ_AnalyzerTM library creation, the standard mixture chromatograms
295 of all the analytes were submitted to the UIE module. This software also has the
296 option of importing new spectra as txt/csv files. This enables the inclusion of spec-
297 tra recorded in the open databases such as MassBank [31] into the local library of
298 NIVA_MZ_AnalyzerTM.

299 3. Results and discussions

300 3.1. Cross platform compatibility

301 The cross platform similarity values of all 7 common analytes were three order
302 of magnitudes larger than the minimum threshold of similarity, see section S5. The
303 largest maximum threshold of similarity at MS¹ was of 0.1627 for Carbamazepine
304 recorded via Waters instrument whereas the smallest value for the maximum sim-
305 ilarity threshold of 0.0072 was observed for Morphine, also recorded with Waters
306 equipment, Figure 1. When looking at the MS², the highest value of 0.071 was
307 registered with Metoprolol via Waters and the lowest value of 0.0071 was observed
308 for Morphine recorded with Waters. The higher variability observed in the Waters
309 instrument compared to the Bruker setup, was attributed to the applied ramp in
310 the collision energy of the MS², while for the Bruker instrument a constant collision
311 energy was applied for the MS² spectra. In the case of cross platform similarity,
312 this parameter ranged from 0.0083 for Citalopram to 0.1105 for Carbamazepine. In
313 overall, for both MS¹ and MS², we observed a slightly lower average cross platform
314 similarities compared to the similarity values within each platform, Figure 1. We
315 interpreted that the observed decrease in the cross platform similarities compared
316 to the maximum threshold similarities were caused by the differences in the collision
317 energies and also the hardware design of the two instrumental setups. However, the
318 non-parametric Kruskal-Wallis test [32] at 95% confidence interval, with p values of
319 0.84 indicated that, in both MS¹ and MS², the observed differences between the cross
320 platform similarity values and the maximum threshold similarities were not statis-
321 tically significant. This implied that the average spectra, for the common analytes,

322 recorded by different instruments are comparable to each other. These results were
323 consistent with the findings of Oberacher et al. [33] and Scheubert et al. [34], which
324 suggested by using an average acquisition condition, an effective spectral weighting
325 function, and dot product the cross platform compatibility can be achieved. This
326 implies that these spectra can be used for suspect and/or retrospective analyses of
327 samples acquired via different instruments. However, a more comprehensive evalua-
328 tion of cross platform compatibility with a large pool of instruments and analytes is
329 necessary for this aim and will be the subject of our near future study.

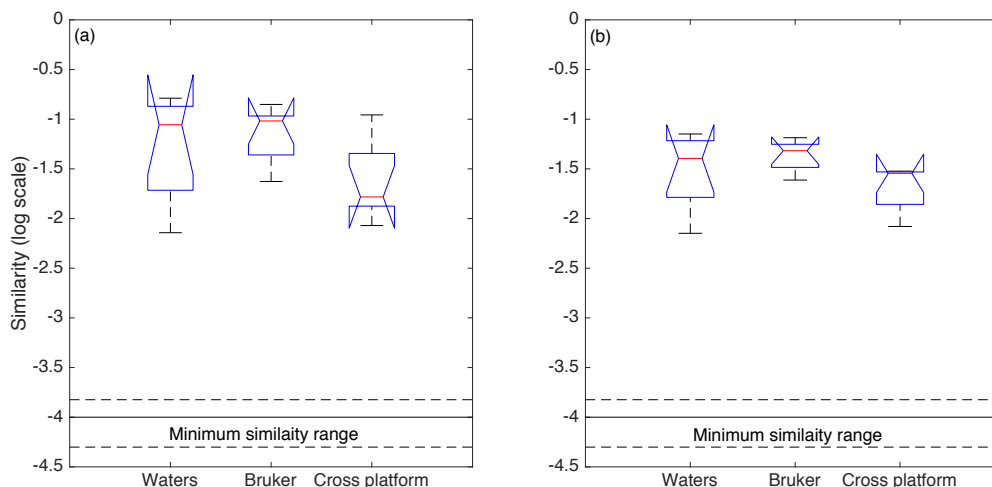


Figure 1: Calculated maximum threshold similarity and cross platform similarity of 7 common analytes at (a) MS¹ and (b) MS². The similarity values are available in Table S4 in Supporting Information. The solid line at the bottom of the figure depicts the averaged minimum threshold of similarity whereas the dashed lines show the variability observed in the minimum threshold of similarity.

330 *3.2. Suspect analysis and analyte identification*

331 Four wastewater samples (i.e. Oslo samples and Athens samples) were analyzed
332 for 89 analytes with NIVA_MZ_AnalyzerTM in suspect analysis mode. The Oslo
333 samples were analyzed in suspect analysis mode with UNIFI whereas the Athens
334 samples were analyzed via TASQ. As a validation tool we analyzed the Oslo samples
335 for Oslo analytes plus the common analytes (i.e. total of 40 analytes out of 89)
336 in target analysis mode whereas for Athens samples we utilized TASQ for Athens
337 analytes plus the common analytes (i.e. total of 39 analytes out of 89) also in target
338 analysis mode. For the ease of discussion we are going to divide our analytes in two
339 categories. The first category, herein referred to as artificial suspect analytes, are
340 the analytes which their presence or absence was confirmed via target analysis in
341 that sample. The second category or true suspects were the analytes that were not
342 analyzed via target analysis in that sample, Tables 1 and S1. We further evaluated
343 our results for false negative, false positive, potential false negative, and potential
344 false positive. A false negative refers to the case where an analyte is not detected
345 via suspect analysis however its presence in the sample is confirmed, employing
346 target analysis. A false positive represents a case where during the suspect analysis
347 a compound is detected while its absence in that sample is confirmed via target
348 analysis. Potential false negative refers to cases of pure suspects, where all the
349 evidence indicates the presence of an analyte in the sample, while the investigated
350 software produces a negative detection. For potential false positive, we refer to a
351 case where a software results in positive detection of an analyte even though all the
352 evidence suggest the contrary. In cases of potential false positives and negatives, the

353 absence or presence of the analyte can not be confirmed due to the lack of retention
354 time information. The presence of both artificial suspects and true suspects, enabled
355 us to simulate a real life case and truly evaluate the limitations of each approach.

356 *3.2.1. Analysis of the artificial suspects*

357 NIVA_MZ_AnalyzerTM confidently detected 9 out of 11 artificial suspect analytes
358 in one of the Oslo samples, while detecting 8 out of 10 in the second sample, Table
359 1. The discrepancy in the number of positive detection in the Oslo samples was
360 caused by the false negative detections of Citalopram in both samples, Morphine
361 in the Oslo1 sample, and Pseudoephedrine in Oslo2 sample. For these compounds
362 NIVA_MZ_AnalyzerTM produced a false negative, due to the fact that the detection
363 requirements (i.e. the six point criterion) by the algorithm were not met, Figure
364 2. The similarity matrix in both low and high collision energies clearly indicated
365 the positive detection of these compounds (e.g. Citalopram, Figure 2). However,
366 the qualifiers selected for them were not present. These false negative cases were
367 caused by high level of background signal, thus ion suppression, which affected the
368 most the selected qualifier ions for these analytes. The effect of the background
369 signal on NIVA_MZ_AnalyzerTM has been discussed in detail in section 3.3. For
370 the Athens samples, NIVA_MZ_AnalyzerTM detected 22 out of 23 target validate
371 analytes in the Athens1 sample whereas it detected 23 out of 24 analytes in the
372 Athens2 sample. For the analytes Irbesartan and Cetirizine, in the Athens sam-
373 ples NIVA_MZ_AnalyzerTM in suspect analysis mode produced a positive detection
374 while TASQ in target analysis mode resulted in their negative detections. We fur-
375 ther interrogated the raw data at the expected location (i.e. the retention time)

376 of these analytes in Athens samples. Our investigation showed that both of these
377 two analytes were present in the sample, Figure 3. For Irbesartan and Cetirizine,
378 we interpret that TASQ faced some difficulties in spectral calibration, that resulted
379 in a mass error of $\sim \pm 3$ mDa. Consequently, TASQ reported these two analytes
380 as negative detections. Once widening the mass error window for Irbesartan and
381 Cetirizine all two analytes were detected both during the suspect and target analy-
382 sis. Therefore, we considered these two analytes as detected and artificial suspect in
383 the Athens samples. NIVA_MZ_AnalyzerTM failed to detect Citalopram, D617, and
384 Caffeine in the Athens samples. In this case similar to Citalopram in Oslo samples,
385 the minimum requirements of the algorithm for positive detection of these two ana-
386 lytes in Athens samples were not met. Therefore, NIVA_MZ_AnalyzerTM resulted in
387 false negatives for Citalopram, D617 and Caffeine in the Athens samples. In overall,
388 the NIVA_MZ_AnalyzerTM produced zero cases of false positives while producing few
389 cases of false negative (i.e. 10 cases out of total analyzed cases of 158). This software
390 showed to be a capable tool for confident detection of suspect analytes in a complex
391 sample. Furthermore, since NIVA_MZ_AnalyzerTM does not need peak picking and
392 it utilizes the whole spectrum rather than few qualifiers, during the suspect analysis,
393 it has a lower potential of false positive detection.

394

395 UNIFI detected 13 out of 11 artificial suspect analytes in the Oslo1 sample
396 whereas it detected 10 out of 10 in the Oslo2 sample, Table 1. We observed 2
397 clear cases of false positive for Hydroxycotinine and Mephedrone in Oslo1 sample.
398 These analytes were detected during the suspect analysis, while were not detected in

399 target analysis mode. UNIFI did not produce any case of false negative. However, it
400 showed to have some difficulties to distinguish analytes such as MDMA and Carba-
401 mazepine in both samples. This software produced multiple potential peaks for these
402 two analytes. Therefore, UNIFI needed retention time information in order to dis-
403 tinguish these potential peaks from the actual peaks of MDMA and Carbamazepine.
404 These difficulties were caused by the complexity of the background signal, which
405 contained the exact mass and more than 2 qualifiers of these two analytes in mul-
406 tiple locations. For the sample Oslo2, UNIFI produced results which were 100% in
407 agreement between the suspect and target analysis. Besides the two false positive
408 cases, UNIFI showed to be a reliable tool in suspect analysis of complex samples.
409 However, depending of the complexity of the samples UNIFI may require additional
410 information and further post processing.

411

412 For the Athens samples, after including the two missed analytes (i.e. Irbe-
413 sartan, and Cetirizine), TASQ in suspect analysis mode, detected 28 out of 25
414 artificial suspect analytes in sample Athens1 and 25 out of 26 in Athens2 sam-
415 ple. For Athens1 sample, TASQ resulted in 3 false positives for analytes D,L,N,O-
416 Didesmethylvenlafaxine, Aliskiren, and Picaridin and 0 cases of false negative. In case
417 of Athens2 sample TASQ produced only 1 case of false negative for Venlafaxine. The
418 cases of false positive when performing suspect analysis with TASQ were attributed
419 to the complexity of the samples. Therefore, other compounds, structurally similar
420 to our analytes, were detected instead of those analytes. Similar to UNIFI, TASQ
421 also detected multiple locations for analytes such as Gabapentin and Carbamazepine,

422 which indicated that this algorithm likewise to UNIFI needed retention time infor-
423 mation for these analytes.

424

425 NIVA_MZ_AnalyzerTM was the only algorithm that did not result in any false
426 positives among the tested software, while producing in total 10 cases of false neg-
427 ative. These false negative cases were mainly caused by the ion suppression due to
428 the sample complexity. The other two approaches (i.e. UNIFI and TASQ) produced
429 both cases of false positive and false negative, which was acceptable considering the
430 complexity of the analyzed samples. These cases of false positive and false nega-
431 tive would have been avoided during post processing where the analysts use their
432 knowledge to increase the level of confidence in the identifications. During this
433 study, we took the necessary precautions to avoid the introduction of background
434 knowledge of the analysts into the final results in order to objectively compare the
435 performances of the investigated algorithms. Our results indicate the importance of
436 full spectral comparison (implemented via NIVA_MZ_AnalyzerTM) rather than em-
437 ploying few qualifiers commonly used during suspect analysis in order to have the
438 maximum level of confidence in the identification.

439 3.2.2. True suspect analysis

440 The NIVA_MZ_AnalyzerTM confidently detected 8 analytes out of 49 true sus-
441 pect analytes (i.e. Athens plus Library analytes) in the Oslo1 sample. In the Oslo2
442 sample this algorithm detected 7 out of 49 true suspect analytes. We analyzed
443 these samples for the same 49 analytes using UNIFI. Except one case (i.e. Eprosar-
444 tan in Oslo2 sample) all the positive detections with NIVA_MZ_AnalyzerTM were

Table 1: The number of positive detection, true positives (i.e. artificial suspect analytes) in each sample processed via three tested approaches.

Software	Samples											
	Oslo1			Oslo2			Athens1			Athens2		
	Pos. detec. ^a	True pos. ^b	Pos. detec. ^a	True pos. ^b	Pos. detec. ^a	True pos. ^b	Pos. detec. ^a	True pos. ^b	Pos. detec. ^a	True pos. ^b	Pos. detec. ^a	True pos. ^b
NIVA_MZ_Analyzer™	9	11	8	10	22	23-25 ^c	23	24-26 ^c				
UNIFI	13	11	10	10	-	-	-	-				
TASQ	-	-	-	-	28	22-25 ^c	25	23-26 ^c				

^aThe number of detected analytes in each sample using a specific software package; ^bThe total number of artificial suspect suspect analytes in each sample; ^cAfter confirming the presence of Irbesartan and Cetirizine in the Athens samples these two analytes were added to the total number of true positives in Athens samples.

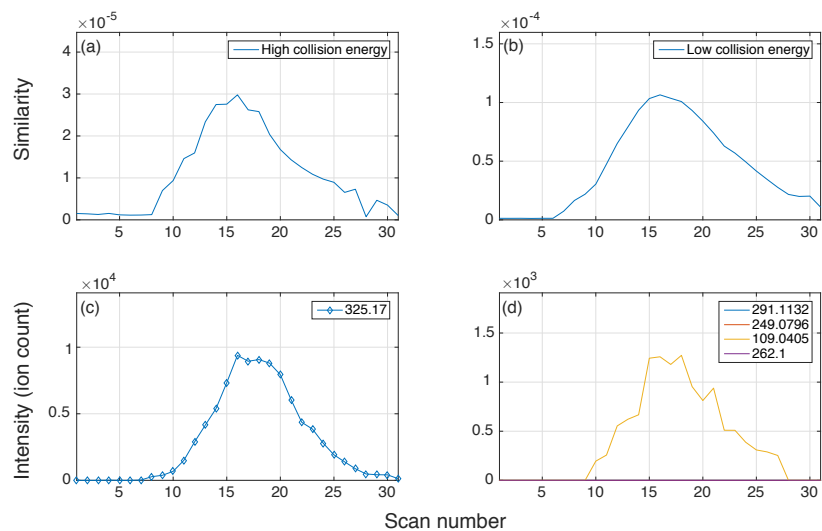


Figure 2: The similarity matrix of Citalopram (a) at MS^2 , (b) at MS^1 , and (c) and (d) the XICs for the exact mass and three qualifiers in Oslo1 sample.

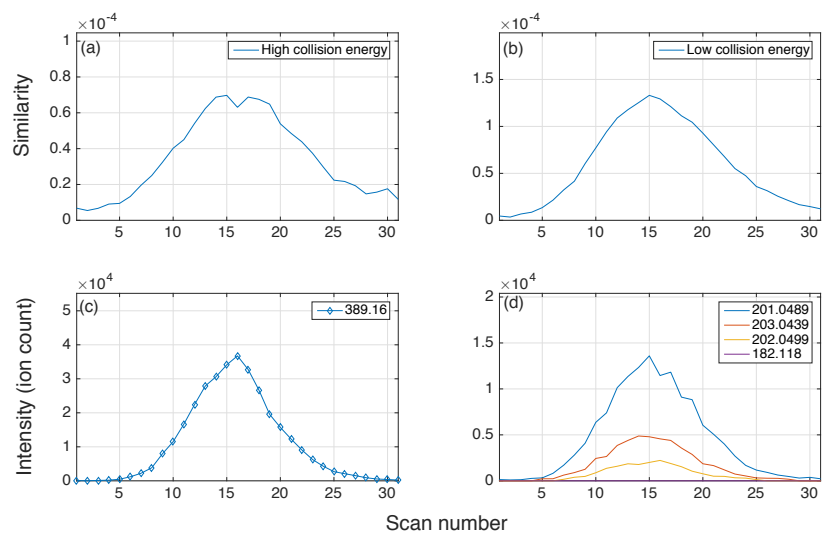


Figure 3: The similarity matrix of Cetirizine (a) at MS^2 , (b) at MS^1 , and (c) and (d) the XICs for the exact mass and three qualifiers in Oslo1 sample.

445 also detected with UNIFI. However, there were several cases of discrepancy where
446 NIVA_MZ_AnalyzerTM resulted in negative detection of an analyte while UNIFI de-
447 tected that analyte in the same sample. We further investigated the observed differ-
448 ences between NIVA_MZ_AnalyzerTM and UNIFI results. More detailed interrogation
449 of the raw data revealed three distinct cases. Case one including analytes Mirtaza-
450 pine, Valsartan, and Caffeine in Oslo1 sample and Valsartan, Caffeine, Ephedrine,
451 and Atenolol acid in Oslo2 sample. For these analytes NIVA_MZ_AnalyzerTM was
452 not able to confirm their presence in the Oslo samples due to lack of 6 positive de-
453 tection criteria. These cases were similar to the Citalopram case in the artificial
454 suspect analysis, Figure 2. Therefore, for these analytes the NIVA_MZ_AnalyzerTM
455 produced potential false negative results. The second case included analytes Clozap-
456 ine, Ecgoninemethylester and para-Methoxy-N-methylamphetamine in Oslo1 sample
457 and Ecgoninemethylester in the Oslo2 sample. By further examining our data, we
458 concluded that UNIFI produced potential false positive results for these analytes,
459 Figure 4. Finally, the third case consisted of Eprosartan. This analyte was detected
460 by NIVA_MZ_AnalyzerTM in Oslo2 sample whereas it was not detected via UNIFI.
461 We scrutinized the Oslo2 dataset and observed that spectral calibration performed
462 via UNIFI for this peak was not successful, which consequently caused its lack of
463 detection in that sample. Therefore it was concluded that this analyte is most likely
464 present in the sample and UNIFI produced a potential false negative result for this
465 analyte, Figure 5.

466

467 NIVA_MZ_AnalyzerTM confidently detected 3 true suspect out of 50 Oslo ana-

468 lytes plus Library analytes in Athens1 sample. This method resulted in 5 positive
469 detections in Athens2 sample. Analysis of the same samples using TASQ resulted in
470 7 positive detections in Athens1 sample and 6 detections for Athens2 sample. By in-
471 terrogating the discrepancy cases between NIVA_MZ_AnalyzerTM and TASQ, we ob-
472 served two different categories of suspect analytes. The first category included Ecgo-
473 ninemethylester, Diclofenac, Hydroxycotinine, and Buprenorphine in Athens1 sample
474 and Ecgoninemethylester, Methylone, and Methylphenidate in Athens2 sample. For
475 these analytes TASQ resulted in positive detections while NIVA_MZ_AnalyzerTM did
476 not detect them. We extracted the signal to noise ratio (S/N) reported for these ana-
477 lytes by TASQ and for all the mentioned analytes in both samples the S/N appeared
478 to be smaller than 9. Furthermore, the results of NIVA_MZ_AnalyzerTM indicated
479 the high level of similarity between the background signal and these analytes' signal,
480 Figure 6. Even though the background knowledge of the wastewater samples may
481 suggest that these analytes are likely to be present in the samples, the data inter-
482 rogation did not provide enough evidence to confirm that. Further investigation is
483 needed in order to draw any type of conclusions regarding these analytes. The second
484 category consisted of analytes Cocaine and 4-MEC in the Athens2 sample. These
485 analytes were not detected with TASQ whereas they were confidently detected via
486 NIVA_MZ_AnalyzerTM. For Cocaine the absolute intensity of the exact mass (i.e.
487 850) was smaller than the threshold of 1250 set by TASQ, therefore this analyte
488 was reported as not detected. When looking at 4-MEC the signal produced by 2
489 qualifiers out of 3 were large and thus they produced broad peaks. The peak picking
490 algorithm in the TASQ failed to detect these broad qualifier peaks therefore TASQ

491 reported 4-MEC as not detected. Based on these evidence we concluded that TASQ
 492 produced a potential false negative for both Cocaine and 4-MEC.

493

494 NIVA_MZ_AnalyzerTM was able to confidently detect the suspect analytes in com-
 495 plex samples. This algorithm did not produce any cases of potential false positives
 496 while resulting in 7 potential false negatives out of 198 cases. The results also indicate
 497 the necessity of full spectral comparison in order to produce high level of confidence
 498 in the identification. In the case of true suspect analysis also NIVA_MZ_AnalyzerTM
 499 appeared to be as successful as commercially available software. In some cases this
 500 algorithm performed far better than both UNIFI and TASQ. UNIFI, resulted in 4
 501 cases of potential false positives whereas TASQ produced 7 cases of inconclusive
 502 outcome. These inconclusive cases are plausible to be potential false positives. In
 503 overall, algorithms such as NIVA_MZ_AnalyzerTM, which perform full spectral anal-
 504 ysis rather than few qualifier inspection are more likely to produce reliable results
 with a high level of confidence.

Table 2: The number of positive detections using the three algorithm investigated in this study.

Softwares	Pure suspects			
	Samples			
	Oslo1	Oslo2	Athens1	Athens2
NIVA_MZ_Analyzer TM	8	7	3	5
UNIFI	15	13	-	-
TASQ	-	-	7	6

505

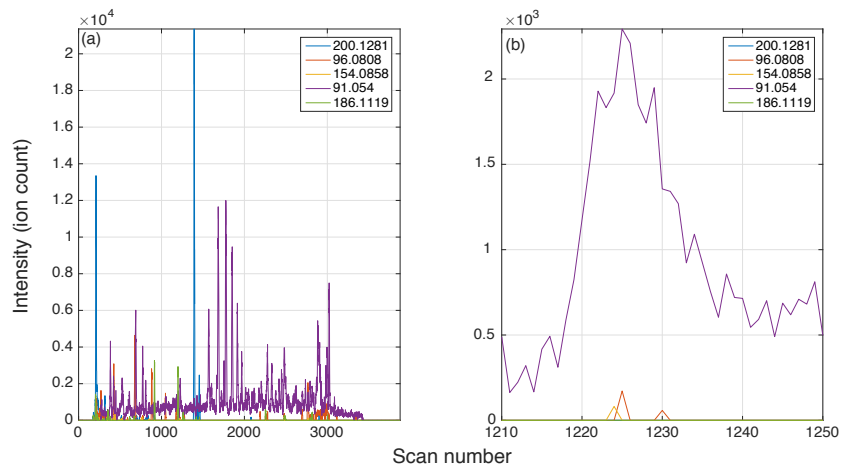


Figure 4: The XICs of the exact mass and 4 relevant ions of Ecgoninemethylester (a) over the whole chromatogram and (b) the zoomed in region based on reported retention time by UNIFI in the Oslo2 sample.

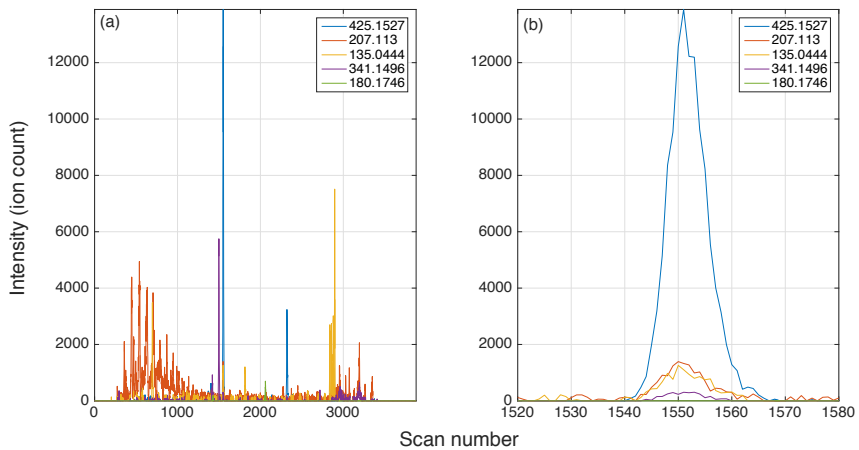


Figure 5: The XICs of the exact mass and 4 relevant ions of Eprosartan (a) over the whole chromatogram and (b) the zoomed in region based on reported retention time by NIVA_MZ_AnalyzerTM in the Oslo1 sample.

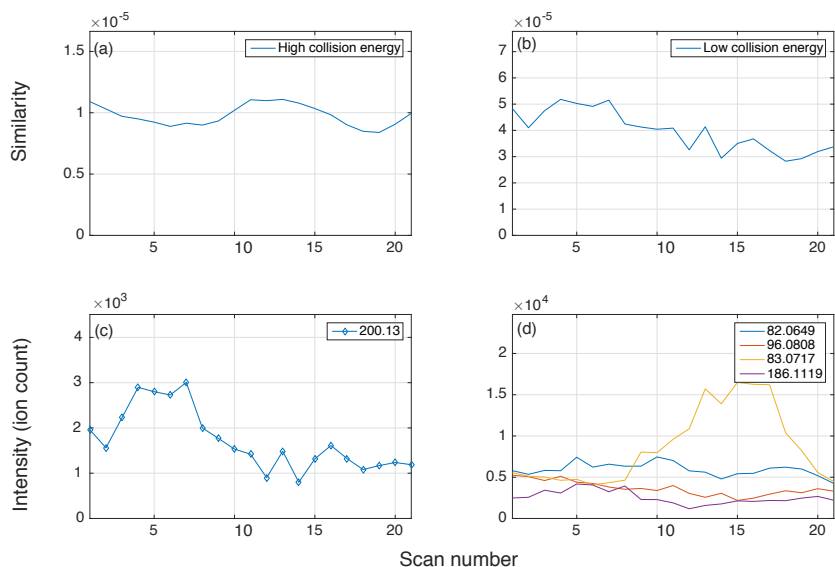


Figure 6: The similarity matrix of Ecgoninemethylester (a) at MS^2 , (b) at MS^1 , and (c) and (d) the XICs for the exact mass and three qualifiers in Athens1 sample.

506 *3.2.3. Rates of false positive and false negative*

507 We compared the 3 algorithms investigated in this study in terms of rates of false
 508 positive and negative, Table 3. The rate of false positives was defined as the ratio of
 509 false positives and false positives plus true negatives whereas the rate of false negative
 510 was defined as the ratio of false negatives and the false negatives plus true positives
 511 [35]. For these calculations we used only the artificial suspect analytes. The rates
 512 of false positive and false negative enabled a direct comparison of these 3 algorithms
 513 and also provided us with indicative levels of confidence in the final identifications
 514 produced by each algorithm.

515

516 NIVA_MZ_AnalyzerTM resulted in the smallest rate of false positive compared to

517 the other two methods. This algorithm produced the highest rate of false negatives
 518 compared to the other algorithms, Table 3. The higher rate of false negative was
 519 attributed to the highly conservative detection criteria used by this algorithm, which
 520 guarantees the highest level of confidence in the identifications. Both UNIFI and
 521 TASQ produced between 0.06 to 0.09 rates of false positives. This again indicated
 522 the importance of full spectral comparison rather than the qualifier approach. The
 523 performances of UNIFI and TASQ were comparable, considering the limited number
 524 of evaluated cases. In overall, the NIVA_MZ_AnalyzerTM appeared to be more robust
 525 compared to the other algorithms when looking at the levels of confidence in the
 526 identifications. Moreover, both UNIFI and TASQ require post processing in order
 527 to avoid cases of false positive, thus a high level of confidence in the identification.

Table 3: The number of true positive, true negative, false positive, false negative, rate of false positive, and rate of false negative calculated for all 3 algorithm investigated in this study.

Algorithm	True +	True -	False +	False -	Rate false +	Rate false -
NIVA_MZ_Analyzer TM	72	88	0	10	0	0.12
UNIFI	21	59	4 ^a	2	0.06	0.09
TASQ	51	29	5 ^b	1-4 ^c	0.15	0.02-0.07 ^d

^aThe number of false positives includes analytes MDMA and Carbamazepine, where the retention time information is necessary for correct detection; ^bThe number of false positives includes analytes Gabapentin and Carbamazepine, where the retention time information is necessary for correct detection; ^cThis number of false negatives includes analytes Irbesartan and Cetirizine, which were not detected due to the calibration issues; ^dThis rate of false negatives includes analytes Irbesartan and Cetirizine, which were not detected due to the calibration issues.

528 3.3. The effect of the sample complexity and concentration on the two stage algorithm

529 We examined the effect of both background complexity and analyte concentration
 530 on the performance of the two stage algorithm. We generated synthetic background

531 signal using real data plus noise. The procedure of the signal generation is explained
532 in detail in Supporting Information, section S6. In our evaluation, we looked at
533 three different levels of complexity in the background signal. These levels consisted
534 of: low complexity, where there were less than $\sim 10\%$ common features between the
535 background signal and the analyte signals; medium complexity, where we observed
536 $\sim 50\%$ similarity between the background signal and the analyte signals; and finally,
537 the high complexity, in which there were $\sim 80\%$ similarity between the background
538 signal and the analyte signals, Figure S3. We carried out the identification employ-
539 ing two stage algorithm for 7 common analytes at 7 signal dilution factors ranging
540 from 1 to 0.01. By applying different signal dilution factors, we observed the signal
541 to background ratio (S/B) varying from 158 to 0.1. The signal was defined as the
542 sum of all intensities in the analyte spectra whereas the background was defined as
543 the averaged sum of all the intensities in the background spectra (See section S6 in
544 SI).

545

546 The two stage algorithm confidently distinguished all 7 analytes from the back-
547 ground signal for all three background levels of complexity at signal dilution factors
548 ≥ 0.25 . The background complexity had a negative effect on the performances of
549 two stage algorithm. For simplicity we are going to focus on two extreme cases
550 Paracetamol and Morphine. These two analytes appeared to be representative for
551 all 7 common analytes. Paracetamol due to its size and structure produced a limited
552 number of ions in both low and high collision energies. Therefore, the two stage
553 algorithm was less affected by complexity of the background signal, Figure 7. For

554 Morphine, the number of low intensity ions (i.e. small fragments) generated, par-
555 ticularly at MS², were larger than Paracetamol. Therefore, the two stage algorithm
556 faced more difficulties in separating the analyte signal from the background. How-
557 ever also in the case of Morphine the two stage algorithm was able to separate the
558 two signals.

559

560 The concentration showed to have a positive effect on the performances of two
561 stage algorithm. At larger concentrations the two stage algorithm was able to sepa-
562 rate the analyte signal from the background signal at all three complexity levels and
563 for all 7 common analytes. At high signal dilution factors (i.e. lower concentrations)
564 with high background complexity the two stage algorithm was not able to confidently
565 detect the analyte signal in the background, Figure 8. The level of dissimilarity for
566 the lowest concentrations of successfully detected analytes via two stage algorithm
567 was reevaluated. The results of this experiment indicate that the two stage algorithm
568 requires at least 10% dissimilarity between the analyte and the background signal in
569 order to be able to separate the two signals from each other. We also evaluated the
570 S/B of 7 common analytes at different concentration levels and background complex-
571 ities. We observed a minimum S/B values of ~ 1 required for two stage algorithm
572 in order to distinguish the analyte signal from the background. For Paracetamol the
573 S/B values of ~ 1 , for both low and high collision energy spectra, were reached at
574 signal dilution factor of 0.1 at the highest background complexity whereas for the
575 low background complexity this S/B ratios were obtained with signal dilution factors
576 0.01. On the other hand, for Morphine, the S/B values of ~ 1 were reached at higher

577 concentrations (i.e. signal dilution factor of 0.25) compared to Paracetamol for all
578 three background complexity levels. Therefore, for an analyte such as Morphine with
579 highly complex background higher concentrations were needed in order for two stage
580 algorithm to separate the analyte signal from the background.

581

582 The two stage algorithm was shown to be able to distinguish the analyte signal
583 from the complex background signal. This algorithm requires S/B values of ≥ 1 for
584 both low and high energy channels in order to be able to distinguish the analyte
585 signal from the background. Furthermore, a minimum of 10% dissimilarity between
586 the analyte signal and the background is needed for the confident identification of
587 an analyte in the complex background.

588 **4. Conclusions**

589 The comparison between an alternative algorithm (i.e. two stage algorithm imple-
590 mented via NIVA_MZ_AnalyzerTM) and two commonly used approaches (i.e. UNIFI
591 and TASQ) was performed via the suspect analysis of 89 analytes in four wastewater
592 influent samples. We observed cross platform similarities larger than 3 orders of
593 magnitude from the minimum threshold of similarity. A slight decrease in the level
594 of similarity was observed in cross platform similarities vs within platform similarity.
595 However, the Kruskal-Wallis test results indicated that the differences observed were
596 not statistically significant. This was indicative of the cross platform applicability
597 of the data produced via different instrument for suspect/retrospective analysis for
598 the instrumental setups evaluated in this study. Further research on this subject is

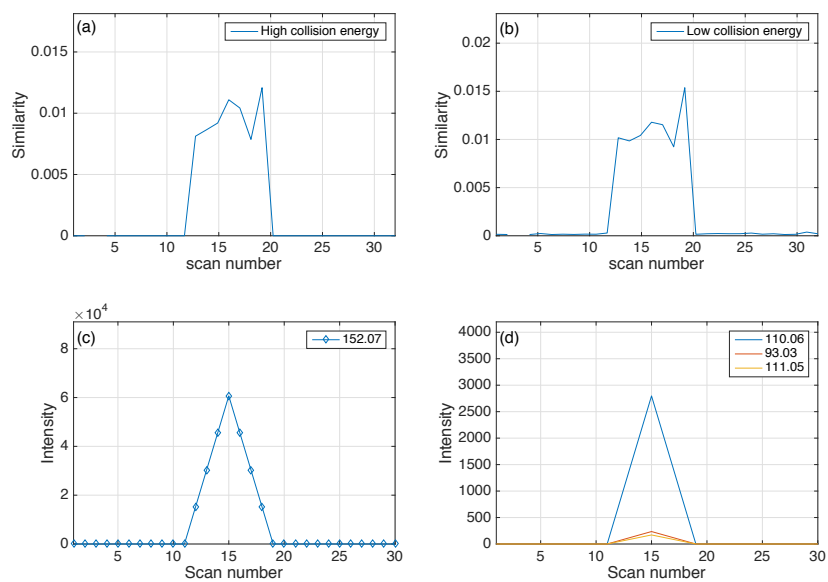


Figure 7: The similarity matrix of Paracetamol (a) at MS^2 , (b) at MS^1 , and (c) and (d) the XICs for the exact mass and three qualifiers. These data are generated at low background complexity and the lowest signal dilution factor (i.e. 1).

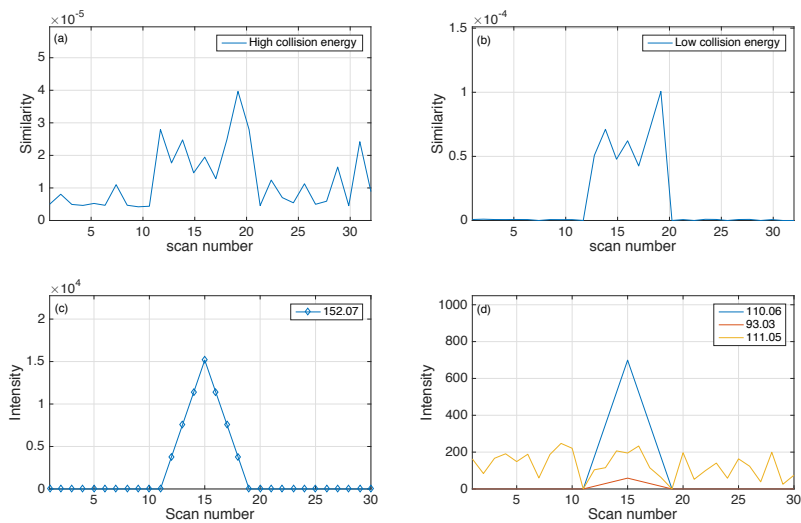


Figure 8: The similarity matrix of Paracetamol (a) at MS^2 , (b) at MS^1 , and (c) and (d) the XICs for the exact mass and three qualifiers. These data are generated at high background complexity and the highest signal dilution factor (i.e. 0.25).

599 needed in order to validate a universal data acquisition procedure for LC-HR-QTOF-
 600 MS data.

601

602 For the artificial suspect analytes, the two stage algorithm did not produced any
 603 cases of false positive while resulting in 10 false negatives out of total 158 artificial
 604 suspect cases. UNIFI and TASQ both produced ~ 5 cases of false positives and
 605 between 0 to 4 cases of false negatives. For true suspects, the two stage algorithm
 606 again did not produced any cases of potential false positives while resulting in 7
 607 cases of potential false negatives. UNIFI for the pure suspects produced 4 cases of
 608 potential false positives and 1 case of potential false negative. TASQ for the pure
 609 suspects produced 7 cases of inconclusive/potential false positive and 2 cases of false

610 negative. Our direct comparison of the three approaches with the means of rate
611 of false positive indicated that the two stage algorithm with a rate of false positive
612 of 0 resulted in the most robust method with the highest level of confidence in the
613 identifications.

614

615 The evaluation of the background complexity effect on the two stage algorithm
616 showed that this algorithm requires at least 10% of dissimilarity between the back-
617 ground and the analyte signal. Three levels of complexity ranging from 20% dissim-
618 ilarity (i.e. the highest level of complexity) to 90% dissimilarity between the analyte
619 and background signal were evaluated. We also evaluated the effect of S/B ratio
620 on the performances of the two stage algorithm. We observed that this algorithm
621 needs ≥ 1 S/B ratios for both MS¹ and MS² in order for the algorithm to produce
622 the highest level of confidence in the identification. For the conventional approaches,
623 the signal to noise ratio was considered an indication of the background complexity.
624 Overall, the two stage algorithm showed to be more affected by the signal suppres-
625 sion compared to the conventional methods, due to its use of the full spectra. These
626 evaluation enabled us to define the boundaries in which the two stage algorithm is
627 capable to produce reliable results.

628

629 The two stage algorithm as well as UNIFI and TASQ were capable of performing
630 suspect analysis in the investigated complex samples. However, contrary to the two
631 stage algorithm both UNIFI and TASQ needed fine tuning of the filters and post
632 processing in order to minimize the odds of false positive detections. With UNIFI

633 and TASQ the background knowledge and the experience of the analyst plays an
634 important role in increasing the level of confidence in the final identification.

635

636 This study showed the cross platform compatibility of the data produced via two
637 tested HR-TOF-MS instruments. This implies that generation of a universal HR
638 spectral database similar to the NIST library is also possible for LC-HR-QTOF-MS
639 data. However, more efforts and future studies are needed in order to establish a
640 standard operational procedure (SOP) for LC-HR-QTOF-MS data acquisition.

641 **5. Acknowledgements**

642 The authors are thankful to Prof. Bert Van Bavel and Dr. Katherine Langford
643 for their help during this project. We are also thankful to the Research Council of
644 Norway for the financial support of this project (RESOLVE, 243720).

645 6. References

- 646 [1] Eva Gorrochategui, Joaquim Jaumot, Sílvia Lacorte, and Romà Tauler. Data
647 analysis strategies for targeted and untargeted LC-MS metabolomic studies:
648 Overview and workflow. *Trends Anal. Chem.*, 82:425–442, 2016.
- 649 [2] Emma L Schymanski, Heinz P Singer, Jaroslav Slobodnik, Ildiko M Ipolyi, Peter
650 Oswald, Martin Krauss, Tobias Schulze, Peter Haglund, Thomas Letzel, Sylvia
651 Grosse, et al. Non-target screening with high-resolution mass spectrometry:
652 critical review using a collaborative trial on water analysis. *Anal. Bioanal.*
653 *Chem.*, 407(21):6237–6255, 2015.
- 654 [3] Emma L Schymanski, Heinz P Singer, Philipp Longree, Martin Loos, Matthias
655 Ruff, Michael A Stravs, Cristina Ripolles Vidal, and Juliane Hollender. Strate-
656 gies to characterize polar organic contamination in wastewater: exploring
657 the capability of high resolution mass spectrometry. *Environ. Sci. Technol.*,
658 48(3):1811–1818, 2014.
- 659 [4] Christoph Moschet, Alessandro Piazzoli, Heinz Singer, and Juliane Hollender.
660 Alleviating the reference standard dilemma using a systematic exact mass sus-
661 pect screening approach with liquid chromatography-high resolution mass spec-
662 trometry. *Anal. Chem.*, 85(21):10312–10320, 2013.
- 663 [5] Pablo Gago-Ferrero, Emma L. Schymanski, Anna A. Bletsou, Reza Aalizadeh,
664 Juliane Hollender, and Nikolaos S. Thomaidis. Extended suspect and non-target
665 strategies to characterize emerging polar organic contaminants in raw wastew-
666 ater with LC-HRMS/MS. *Environ. Sci. Technol.*, 49(20):12333–12341, 2015.

- 667 [6] Aurea C Chiaia-Hernandez, Emma L Schymanski, Praveen Kumar, Heinz P
668 Singer, and Juliane Hollender. Suspect and nontarget screening approaches to
669 identify organic contaminant records in lake sediments. *Anal. Bioanal. Chem.*,
670 406(28):7323–7335, 2014.
- 671 [7] Bin Zhou, Jun Feng Xiao, Leepika Tuli, and Habtom W Resson. LC-MS-based
672 metabolomics. *Molecular BioSystems*, 8(2):470–481, 2012.
- 673 [8] Lunzhao Yi, Naiping Dong, Yonghuan Yun, Baichuan Deng, Dabing Ren, Shao
674 Liu, and Yizeng Liang. Chemometric methods in data processing of mass
675 spectrometry-based metabolomics: A review. *Anal. Chem. acta*, 914:17–34,
676 2016.
- 677 [9] Mikko Katajamaa, Jarkko Miettinen, and Matej Orešič. Mzmine: toolbox for
678 processing and visualization of mass spectrometry based molecular profile data.
679 *Bioinformatics*, 22(5):634–636, 2006.
- 680 [10] Colin A Smith, Elizabeth J Want, Grace O’Maille, Ruben Abagyan, and Gary
681 Siuzdak. XCMS: processing mass spectrometry data for metabolite profiling
682 using nonlinear peak alignment, matching, and identification. *Anal. chem.*,
683 78(3):779–787, 2006.
- 684 [11] Marc Sturm, Andreas Bertsch, Clemens Gröpl, Andreas Hildebrandt, Rene Hus-
685 song, Eva Lange, Nico Pfeifer, Ole Schulz-Trieglaff, Alexandra Zerck, Knut Rein-
686 ert, et al. Openms—an open-source software framework for mass spectrometry.
687 *BMC bioinformatics*, 9(1):163, 2008.

- 688 [12] Jianqiu Zhang, Elias Gonzalez, Travis Hestilow, William Haskins, and Yufei
689 Huang. Review of peak detection algorithms in liquid-chromatography-mass
690 spectrometry. *Current genomics*, 10(6):388–401, 2009.
- 691 [13] Jamie B Coble and Carlos G Fraga. Comparative evaluation of preprocessing
692 freeware on chromatography/mass spectrometry data for signature discovery. *J.*
693 *Chromatogr. A*, 1358:155–164, 2014.
- 694 [14] Gunnar Libiseller, Michaela Dvorzak, Ulrike Kleb, Edgar Gander, Tobias Eisen-
695 berg, Frank Madeo, Steffen Neumann, Gert Trausinger, Frank Sinner, Thomas
696 Pieber, et al. Ipo: a tool for automated optimization of xcms parameters. *BMC*
697 *bioinformatics*, 16(1):1, 2015.
- 698 [15] Ralf Tautenhahn, Christoph Böttcher, and Steffen Neumann. Highly sensitive
699 feature detection for high resolution LC/MS. *Bioinformatics*, 9(1):1, 2008.
- 700 [16] Frans M Van Der Kloet, Ivana Bobeldijk, Elwin R Verheij, and Renger H
701 Jellema. Analytical error reduction using single point calibration for accu-
702 rate and precise metabolomic phenotyping. *J. Proteome. Res.*, 8(11):5132–5141,
703 2009.
- 704 [17] Saer Samanipour, Katherine Langford, Malcolm J Reid, and Kevin V Thomas.
705 A two stage algorithm for target and suspect analysis of produced water via gas
706 chromatography coupled with high resolution time of flight mass spectrometry.
707 *J. Chromatogr. A*, 1463:153–161, September 2016.
- 708 [18] Jose A. Baz-Lomba, Malcolm J. Reid, and Kevin V. Thomas. Target and sus-

- 709 pect screening of psychoactive substances in sewage-based samples by UHPLC-
710 QTOF. *Anal. Chem. acta*, 914:81–90, 2016.
- 711 [19] Jose Antonio Baz-Lomba, Stefania Salvatore, Emma Gracia-Lor, Richard Bade,
712 Sara Castiglioni, Erika Castrignanò, Ana Causanilles, Felix Hernandez, Barbara
713 Kasprzyk-Hordern, Juliet Kinyua, et al. Comparison of pharmaceutical, illicit
714 drug, alcohol, nicotine and caffeine levels in wastewater with sale, seizure and
715 consumption data for 8 european cities. *BMC public health*, 16(1):1035, 2016.
- 716 [20] Grung Merete, S. Heimstad Eldbjorg, Moe Morten, Schlabach Martin, Svenson
717 Anders, Thomas Kevin, and Woldegiorgis Andreas. Human and veterinary phar-
718 maceuticals, narcotics, and personal care products in the environment. Technical
719 report, IVL, NILU and NIVA, 2007.
- 720 [21] Nikiforos A Alygizakis, Pablo Gago-Ferrero, Viola L Borova, Alexandra Pavli-
721 dou, Ioannis Hatzianestis, and Nikolaos S Thomaidis. Occurrence and spatial
722 distribution of 158 pharmaceuticals, drugs of abuse and related metabolites in
723 offshore seawater. *Sci. Total Environ.*, 541:1097–1105, 2016.
- 724 [22] Susanne Kern, Kathrin Fenner, Heinz P Singer, René P Schwarzenbach, and
725 Juliane Hollender. Identification of transformation products of organic contam-
726 inants in natural waters by computer-aided prediction and high-resolution mass
727 spectrometry. *Environ. Sci. Technol.*, 43(18):7039–7046, 2009.
- 728 [23] Martin Krauss, Heinz Singer, and Juliane Hollender. LC–high resolution MS in
729 environmental analysis: from target screening to the identification of unknowns.
730 *Anal. Bioanal. Chem.*, 397(3):943–951, 2010.

- 731 [24] Petros Dimitriou-Christidis, Alex Bonvin, Saer Samanipour, Juliane Hollender,
732 Rebecca Rutler, Jimmy Westphale, Jonas Gros, and J. Samuel Arey. GC \times GC
733 quantification of priority and emerging nonpolar organic micropollutants in all
734 types of wastewater streams: Analysis methodology, chemical occurrence, and
735 partitioning. *Environ. Sci. Technol.*, 49:7914–7925, 2015.
- 736 [25] MATLAB User’s Guide. The mathworks. *Inc., Natick, MA*, 5, 1998.
- 737 [26] Pan Du, Warren A Kibbe, and Simon M Lin. Improved peak detection in
738 mass spectrum by incorporating continuous wavelet transform-based pattern
739 matching. *Bioinformatics*, 22(17):2059–2065, 2006.
- 740 [27] Anna A Bletsou, Junho Jeon, Juliane Hollender, Eleni Archontaki, and Niko-
741 laos S Thomaidis. Targeted and non-targeted liquid chromatography-mass spec-
742 trometric workflows for identification of transformation products of emerging
743 pollutants in the aquatic environment. *Trends Anal. Chem.*, 66:32–44, 2015.
- 744 [28] Bruker corporation. TASQTM software ver1.1 - Target Analysis for
745 Screening and Quantitation (TASQ). [www.bruker.com/products/
746 mass-spectrometry-and-separations/ms-software/tasq/tasq-software.
747 html](http://www.bruker.com/products/mass-spectrometry-and-separations/ms-software/tasq/tasq-software.html), August 2016.
- 748 [29] R Mistrik, J Lutisan, Y Huang, M Suchy, J Wang, and M Raab. mzCloud: A
749 key conceptual shift to understand ‘who’s who’ in untargeted metabolomics. In
750 *Metabolomics Society 2013 Conference, Glasgow, July*, pages 1–4, 2013.
- 751 [30] Maria Vinaixa, Emma L Schymanski, Steffen Neumann, Miriam Navarro,

- 752 Reza M Salek, and Oscar Yanes. Mass spectral databases for LC/MS-and
753 GC/MS-based metabolomics: State of the field and future prospects. *TrAC*
754 *Trends Anal. Chem.*, 78:23–35, 2016.
- 755 [31] Hisayuki Horai, Masanori Arita, Shigehiko Kanaya, Yoshito Nihei, Tasuku
756 Ikeda, Kazuhiro Suwa, Yuya Ojima, Kenichi Tanaka, Satoshi Tanaka, Ken
757 Aoshima, et al. Massbank: a public repository for sharing mass spectral data
758 for life sciences. *J. Mass. Spectrom.*, 45(7):703–714, 2010.
- 759 [32] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance
760 analysis. *J. Amer. Statist. Assoc.*, 47(260):583–621, 1952.
- 761 [33] Herbert Oberacher, Marion Pavlic, Kathrin Libiseller, Birthe Schubert, Michael
762 Sulyok, Rainer Schuhmacher, Edina Csaszar, and Harald C Köfeler. On the
763 inter-instrument and inter-laboratory transferability of a tandem mass spectral
764 reference library: 1. results of an austrian multicenter study. *J. Mass. Spectrom.*,
765 44(4):485–493, 2009.
- 766 [34] Kerstin Scheubert, Franziska Hufsky, and Sebastian Böcker. Computational
767 mass spectrometry for small molecules. *J. Cheminform.*, 5(1):1, 2013.
- 768 [35] Neil A Macmillan and C Douglas Creelman. *Detection theory: A user’s guide.*
769 Psychology press, 2004.