2018-09-26

Assessing sample extraction e ciencies for the analysis of complex unresolved mixtures of organic pollutants: A comprehensive non-target approach

Rowland, SJ

Analytica Chimica Acta
Elsevier Masson

Assessing sample extraction efficiencies for the analysis of complex unresolved mixtures of organic pollutants: A comprehensive non-target approach

Saer Samanipour,*,† Jose A. Baz-Lomba,† Malcolm J. Reid,† Elena Ciceri,‡ Steven Rowland,¶ Per Nilsson,§ and Kevin V. Thomas†,‖

†*Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway*

‡*Thermo Fisher Scientific, Strada Rivoltana 20090 Rodano, Milan, Italy*

¶*School of Geography, Earth and Environmental Sciences (Faculty of Science & Engineering) Drake Circus, Plymouth, Devon, PL4 8AA*

§*Thermo Fisher Scientific, Telefonvägen 30, 126 26 Hägersten, Sweden*

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

E-mail: saer.samanipour@niva.no

Abstract

The comprehensive extraction recovery assessment of organic analytes from complex samples such as oil field produced water (PW) is a challenging task. A targeted approach is usually used for recovery and determination of compounds in these types of analysis. Here we suggest a more comprehensive and less biased approach for the extraction recovery assessment of complex samples. This method combines conventional

targeted analysis with a non-targeted approach to evaluate the extraction recovery of complex mixtures. Three generic extraction methods: liquid-liquid extraction (Lq), and solid phase extraction using HLB cartridges (HLB), and the combination of ENV+ and C8 (ENV) cartridges, were selected for evaluation. PW was divided into three parts: non-spiked, spiked level 1, and spiked level 2 for analysis. The spiked samples were used for targeted evaluation of extraction recoveries of 65 added target analytes comprising alkanes, phenols, and polycyclic aromatic hydrocarbons, producing absolute recoveries. The non-spiked sample was used for the non-targeted approach, which used a combination of the F-ratio method and apex detection algorithm. Targeted analysis showed that the use of ENV cartridges and the Lq method performed better than use of HLB cartridges, producing absolute recoveries of 53.1 $\pm$ 15.2 for ENV and 46.8 $\pm$ 13.2 for Lq versus 19.7 $\pm$ 6.7 for HLB. These two methods appeared to produce statistically similar results for recoveries of analytes, whereas they were both different from the produced recoveries via the HLB method. The non-targeted approach captured unique features that were specific to each extraction method. This approach generated 26 unique features (mass spectral ions), which were significantly different between samples and were relevant in differentiating each extract from each method. Using a combination of these targeted and non-targeted methods we evaluated the extraction recovery of the three extraction methods for analysis of PW.

Introduction

Comprehensive extraction recovery assessments of complex mixtures of organic analytes are extremely difficult. This is caused mainly by the complexity of the sample and lack of knowledge regarding the chemical constituents of the sample. Consequently, a generic/wide range extraction method is typically employed for the analysis of complex mixtures such as produced water (PW; reviewed by Oetjen[1]). Often, different extraction methods are tested on a small number of potential target analytes (compared to the number of chemicals in

a complex mixture) in order to define an optimized extraction method.[1,2] This approach assumes that the fate and behavior of each chemical constituent in the complex mixture can be linearly extrapolated by the behavior of the target analytes and that there are no interactions between different chemicals. Such an approach is perhaps questionable, for example, when an examination of PW for naphthenic acids is made, since these compounds also behave as surfactants. Another method used for the extraction recovery assessment of complex mixtures is the gravimetric approach.[1,3] This method focuses on the total non-volatile extractable material. In this case if the amount of a certain chemical in the sample is smaller than the experimental error (e.g. $\pm$ 10%) then it is impossible to capture any mass loss for that chemical caused by different extraction methods. Therefore, both mentioned methods are not applicable to comprehensively evaluate the recovery of different extraction methods when dealing with complex mixtures such as PW.

PW is one of the largest streams of treated industrial wastewater in the world[4] and its discharge into the marine environment is of ecological relevance. For example from Norwegian off shore activities PW volumes are 140 mil m^3 y^{-1}.[5] PW is a complex mixture containing a diverse range of chemical constituents.[1,6–8] Organic compounds in PW, typically vary from oil droplets to large organic acids.[6–8] Thus, PWs exhibit a wide range of chemical and physical properties, fate and behaviors. As a consequence of this chemical diversity and the fact that not all of its chemical constituents are known, extraction of PW typically reveals complex mixtures that are largely unresolved by typically used techniques (e.g. unit mass GC-MS).[9–11]

High resolution mass spectrometry coupled with different chromatographic technologies (gas and/or liquid chromatography) has shown great potential in partially resolving the unresolved complex mixture (UCM).[12–15] However, when dealing with UCMs, these analytical techniques are not capable of comprehensively characterize the analyzed samples.[14] Conse-

quently, chemometric tools such as principal component analysis (PCA), F-ratio, and N-way partial least-squares in combination with HRMS are usually employed to tackle the complexity of these UCMs.[15–18]

The combination of F-ratio method and the apex detection algorithm has been shown to be a powerful tool when dealing with complex environmental samples, including petroleum related matrix.[17,20] F-ratio is a parametric supervised method, which uses the ratio of the between-groups variability and within each group variability to define the significance of each variable.[19,20] Therefore, it identifies the features in the samples which are statistically significant, while the apex detection algorithm reduces the redundancy in those features by grouping them as unique statistically significant feature. PW was selected as the test/validation matrix for the applicability of this approach in comprehensive recovery assessment of complex mixtures due to its complexity.

The aim of the present study was to use the F-ratio method to comprehensively assess the extraction recovery of three generic (i.e. wide range of chemical and physical property) extraction methods for PW. We employed three extraction methods: liquid-liquid extraction (Lq), HLB cartridges (HLB), and the combination of ENV+ and C8 cartridges(ENV) for an applicability proof of concept. These methods have been widely used for recovering complex mixtures of analytes from matrices including PW.[21–26] We employed a combination of the conventional targeted and the alternative non-targeted analysis for a comprehensive recovery assessments. PW was divided into three categories: non-spiked, spiked level 1, and spiked level 2. For the targeted approach we used a spike solution consisting of a mixture of 65 target analytes that were added into the PW at two different concentrations (i.e. spiked level 1 and spiked level 2). The concentration differences between the two spike levels were used to calculate the absolute recoveries of each target analyte. For the non-targeted approach, we used the non-spiked PW. We employed the null-distribution in order to define the threshold

of false positive detection. Finally, we calculated the relative recovery of unique features based on the average intensity of those features. This study was a proof of concept for the applicability of the suggested approach in comprehensive recovery assessment of complex unresolved mixtures of organic analytes.

Experimental Methods

Sample Preparation and Extraction

PW (20L) was obtained from the Heidrun oil platform[27] in the Halten bank off the coast of mid-Norway during February 2017. PW was subdivided into 27 aliquots each of 400 mL. These aliquots were divided into three categories: non-spiked, spiked level 1 and spiked level 2, thus 9 samples in each category (Figure 1). We added a predefined volume of a standard mixture solution to the spiked samples (i.e. spiked level 1 and spiked level 2) in order to reach a certain concentration for each added component of the mixture. The standard mix solution consisted of a mixture 29 alkanes (Als) from C10-C33 at 8 μg mL^{-1} each, 19 alkylated phenols (ALPs) at 10 μg mL^{-1} each, and 16 polycyclic aromatic hydrocarbons (PAHs) at 2 μg mL^{-1} each. The spiked level 1 samples (i.e. 9 out of 27) were spiked with 50 μL of standard mix solution resulting in addition of 0.4 μg of Als, 0.5 μg of ALPs, and 0.1 μg of PAHs whereas spiked level 2 samples were spiked with 100 μL of standard mix solution resulting in addition of 0.8 μg of Als, 1 μg of ALPs, and 0.2 μg of PAHs. The non-spiked samples were used for non-targeted recovery assessment while the spiked samples were employed for the targeted workflow. Detailed information regarding the standard mixtures and suppliers is provided in the Supporting Information, Section S1.1 and Table S1.

Each spiked level sample group was extracted using one of three different extraction methods: liquid-liquid extraction (Lq), HLB cartridges, or the combination of ENV+ and C8 cartridges (ENV), each in triplicates, Figure 1. The Lq method resulted in recovering

a dichloromethane extract of acidified PW (pH 2). This method is the official method recommended by the Norwegian Oil and Gas for extraction of PW.[25] On the other hand, use of the HLB cartridge is a solid phase extraction (SPE) approach, where the solid phase is a universal polymeric reverse phase sorbent for extraction of acidic, basic and neutral compounds in different water-based matrices. This method has been widely used for analysis of wastewater samples.[21–24] ENV+ is another SPE cartridge with a non-polar crosslinked hydroxylated polystyrene-divinylbenzene solid phase, reportedly adequate for extraction of polar and semi-polar compounds from complex aqueous samples.[26] The combination of ENV+ and the reversed phase C8 cartridges enables extraction of a wide range of chemicals with polarity varying from non-polar to polar. This method has been successfully used for extraction of PW, previously.[26] More detailed information regarding the extraction procedures is provided in the Section S1.2 of the Supporting Information. The three tested methods all are considered to be generic extraction methods, which implies that they are supposed to extract a large number of chemical constituents with a wide range of chemical and physical properties in the PW.

For the quality control/assurance of the analysis, we took the following steps during our extractions. For application of each extraction method at a specific spiked level, a procedural blank was generated, Figure 1. These procedural blanks were extracts of either the unloaded cartridges or the glassware used for Lq method. All the glassware used during the extractions and analyses was oven baked at 450 °C over-night. Additionally, all the final extracts were spiked with 50 ng of diazepam-d5 as injection standard in order to monitor the performance of the instrumentation.

Instrumental Conditions and Analysis

The final extracts of non-spiked samples and all the blanks were analyzed via Thermo Scientific™ QExactive™ GC Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Ther-

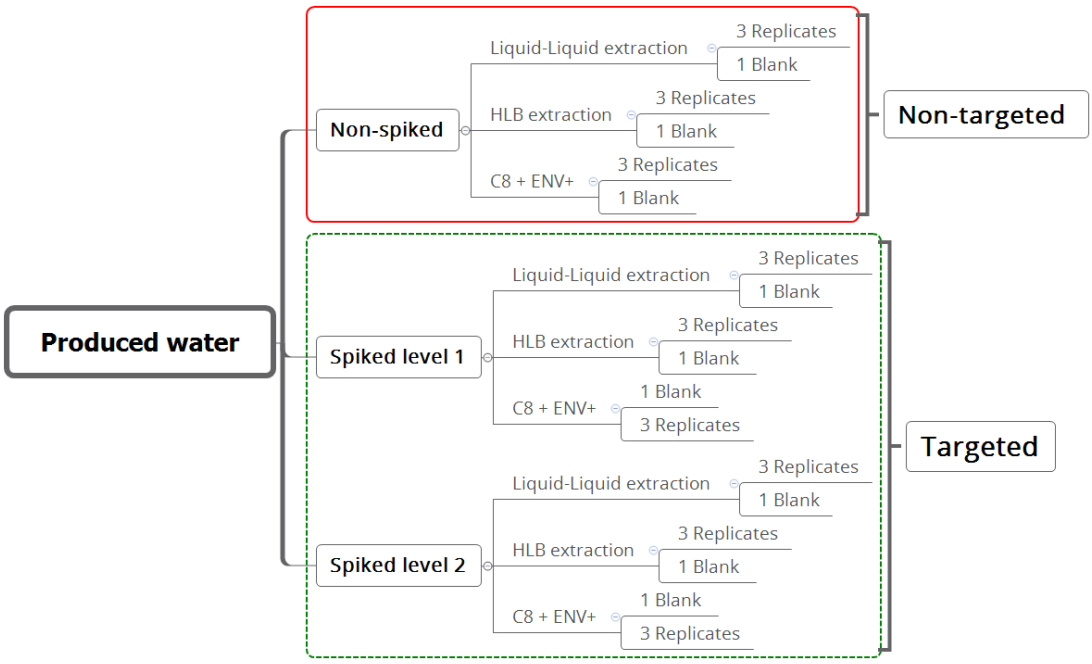


Figure 1: Schematic of the design of the experiment employed in this study depicting the extraction methods, number of replicates, number of spiking levels and data processing approach.

moFisher Scientific, USA) with an electron impact ionization source (EI), hereafter referred to as GC-Orbi. One μL of each extract was injected in splitless mode at 320 °C of inlet temperature. The samples were separated on a 30 m $\times$ 0.25 mm $\times$ 0.25 μm TraceGOLD (TG-5MS) by ThermoFisher Scientific, USA. We employed Thermo Scientific^{TM} TraceFinder^{TM} software (ThermoFisher Scientific, USA) for the data acquisition of the non-spiked samples.

The extracts of spiked levels 1 and 2 samples as well as all the blanks were analyzed employing GC coupled to a high resolution time of flight mass spectrometery (GC-HR-TOFMS; GCT Premier, Waters, USA) equipped with EI source. The samples were examined using a DB-5 column (30 m $\times$ 0.25 mm $\times$ 0.25 μm, Agilent) with an injection volume of 1 μL. The TOFMS was operated with a sampling frequency of 2 Hz between 50 and 650 Da with a resolution of 9000 at half width full range. The chromatograms of these samples were acquired via MassLynx^{TM} (Waters, USA). These settings were optimized previously for

analysis of PW extracts.[28] The details regarding the temperature program used for these separations are provided in the Section S1.3 of the Supporting Information.

Target Analysis and Absolute Recovery Assessment

Target screening was employed for the analysis of the spiked level 1 and 2 samples. Details of the detection and quantification procedure are provided elsewhere.[28] In brief, we used the retention time, accurate mass of the parent ion and the accurate masses of two fragments for confident identification of the target analytes while using a five point external standard calibration curve with three replicates at each level for the quantification of the target analytes. The differences in the average concentration of the analytes between spiked level 2 and spiked level 1 were used for the absolute recovery calculations. Throughout this document we refer to the recoveries calculated via target analysis as absolute recoveries. It should be noted that the analytes which produced a negative or zero absolute recoveries were considered to have a recovery of zero.

Data Processing for Non-targeted Recovery Assessment

The raw chromatograms of the non-spiked samples were converted to mzXML format employing the MSConvert package implemented via ProteoWizard.[29] The converted data files were imported into Matlab (R2015b)[30] for further processing. During the non-targeted data processing the imported data went through five consecutive steps: 1) data binning, 2) retention alignment, 3) F-ratio calculation, 4) null distribution, and finally 5) Apex detection (Figure S1). The F-ratio method, being a parametric test, assumes normal distribution of the tested dataset. Typically, the data produced via LC-MS and/or GC-MS are more than 65% normally distributed, which implies the adequacy of a parametric method for the analysis.[31] This is particularly the case for the raw LC-MS and GC-MS data due to inherent nature of the raw data, which consist of a combination of gaussian peaks for analytical signal and noise. Therefore, the F-ratio method can be applied to these datasets. We selected a

very large F-ratio threshold with a very small probability of false positive detection of 0.01%. The reason behind this choice of F-ratio value was the fact that this study is only a proof of concept, and therefore, we preferred to focus on a limited number (i.e. sub-sample) of the unique statistically relevant features rather than all of them. This workflow has been shown to be able to capture the statistically meaningful differences between different sample sets.[17] The details of all the steps in the non-targeted workflow is available in the Section S2 of the Supporting Information.

For the non-targeted recovery assessment, hereafter referred to as relative recoveries, the average signal of the method with highest intensity for a certain feature is assumed to be the total extractable material for that feature. Therefore, the ratio of the average signal of a certain feature for all the extraction methods and the total extractable material could be considered the relative recovery of that feature via that extraction method. In Eq. 1, Rec_{Rel} represents the relative recovery, $\hat{S}_{i,j}$ represents the average signal of i^{th} feature and j^{th} extraction method, and $\hat{S}_{i,total}$ represents the total extractable material for i^{th} feature. Using this approach we were able to capture the relative amount of signal lost for a feature due to a specific extraction method.

$$Rec_{Rel} = 100 \times \frac{\hat{S}_{i,j}}{\hat{S}_{i,total}} \tag{1}$$

Computations

All the mentioned data processing steps were performed via Matlab, employing a Windows 7 Professional version (Microsoft Inc, USA) workstation computer with 12 CPUs and 128 GB of memory.

Results and discussion

We comprehensively evaluated the extraction recovery of a complex unresolved mixture, such as PW, via the combination of targeted and non-targeted analysis. Through the target screening we examined the absolute recovery of 65 analytes with three different extraction methods. This was carried out by spiking the PW with a standard mixture at two concentration levels. The concentration differences between the two spike levels were used to calculate the absolute recovery of each target analyte. Additionally, as a quality assurance step we evaluated the concentration of the 65 target analytes in the blanks. For all 65 target analytes the sample concentrations were at least 10 times higher than their blank concentrations.The non-targeted approach, on the other hand, was used to capture the statistically meaningful features in the samples which differentiated each extraction method from the others. We used the F-ratio method in order to select the relevant features in each sample.[17,32,33] The F-ratio method was combined with the null distribution approach to calculate the probability of false positive detection for each F-ratio.[17,20] During the F-ratio analysis, the blanks for each extraction method (i.e. the non-spiked and the two spike levels) were grouped together as triplicates. These blank triplicates were included in the dataset used for F-ratio analysis as separate groups. This procedure enabled us to assure that the finally selected features are unique to the samples. This study is a proof of concept for the applicability of this approach to comprehensively assess the extraction recovery of unresolved complex mixtures, particularly for non-targeted structural elucidation and/or retrospective analysis.

Targeted Recovery Assessment

The ENV method resulted in the largest number of analytes (i.e. 48 out of 65; 74%) with an absolute recovery larger than zero whereas the HLB method produced the smallest number of positive recovery analytes, 34 out of 65 (52%), Table 1. A similar trend was observed for the average absolute recovery of each extraction method across all three chemical families

(Table 1 and Figure 2). The ENV method was able to extract Als from dodecane to octacosane while the Lq method was more successful in extraction of smaller Als such as decane, Figure S3. In case of ENV method the C8 sorbant had a similar level of affinity towards the Als with different molecular size. Therefore, the higher volatility of these smaller Als compared to the larger ones caused lower recoveries for those analytes. For the Lq method the observed trend was attributed to the higher solubility of smaller Als in the DCM compared to the larger analytes. For these analytes (i.e. Als) the HLB method was less successful than both ENV and Lq methods in extracting the small Als and n-pentadecane was the smallest extracted Al. consequently, for the larger Als, this method fared better than Lq method while performing in a similar way to the ENV method. For ALPs, similarly to the Als, the ENV method extracted the largest number of target analytes (i.e. 13) when compared to the other two methods, Table 1. We were not able to find a consistent trend between the molecular size or hydrophobicity of target analytes and their absolute recoveries. However, all three methods appeared to be more successful in extraction of smaller ALPs (Figure S4). For PAHs, the ENV and Lq methods were able to produce positive recoveries for all 16 target analytes whereas the HLB method was only able to extract 12 analytes out of 16 (Table 1 and Figures 2 and S5). Overall, the ENV and Lq methods performed better than the HLB method based on the observed number of analytes with positive recoveries.

Regarding the absolute recoveries, the ENV and Lq methods with average absolute recoveries of 53.1 $\pm$ 15.2 for ENV and 46.8 $\pm$ 13.2 for Lq performed better than the HLB method with an average absolute recovery of 19.7 $\pm$ 6.7 (Table 1 and Figure 2). The ENV method with an observed within replicates' variability of 59% appeared to be the most stable extraction method compared to HLB method with 85% observed variability and Lq with 198% observed variability (Figures S3, S4 and S5). The Lq method includes more manual steps than the SPE methods. Both ENV and HLB methods showed more uniform recoveries (i.e. closer to the average recovery) across all the target analytes compared to the Lq

method, whereas the Lq method resulted in larger levels of variability in the recoveries as a function of analyte molecular size and DCM solubility (e.g. Als, Figure S3). In terms of absolute recoveries, the ENV and Lq methods performed in a similar way for all three chemical families while the HLB method fared the worst.

The methods ENV and Lq were not statistically distinguishable when looking at all 65 target analytes while they both appeared to be different from the HLB method (Kruskal-Wallis test[34] p value < 0.01). We used the non-parametric Kruskal-Wallis test[34] to differentiate the investigated extraction methods from each other. The observed result of the statistical test was in agreement with the observed trends of recoveries for different chemical families and extraction methods.

Table 1: Lists the number of analytes with positive absolute recoveries as well as the average absolute recoveries for each extraction method and chemical family.

Number of chemicals with positive recoveries[a]			
	Extraction methods		
Chemical family	ENV	HLB	Lq
Al[b]	19	15	19
ALP[c]	13	7	9
PAH[d]	16	12	16
Total	48	34	44

Average absolute recoveries[a]			
	Extraction methods		
Chemical family	ENV	HLB	Lq
Al	52.4±10.2	17.1±7.0	50.0±16.2
ALP	41.1±17.3	14.8±6.4	37.9±6.9
PAH	63.5±17.4	26.1±5.7	48.1±12.0
Total	53.1±15.2	19.7±6.7	46.8±13.2

[a] This parameter was calculated using only the anaytes with positive recoveries; [b] The total number of alkanes (Als) in this study was 29; [c] The total number of investigated alkylated phenols (ALPs) was 19; and [d] The total number of PAHs in this study was 16 compounds.

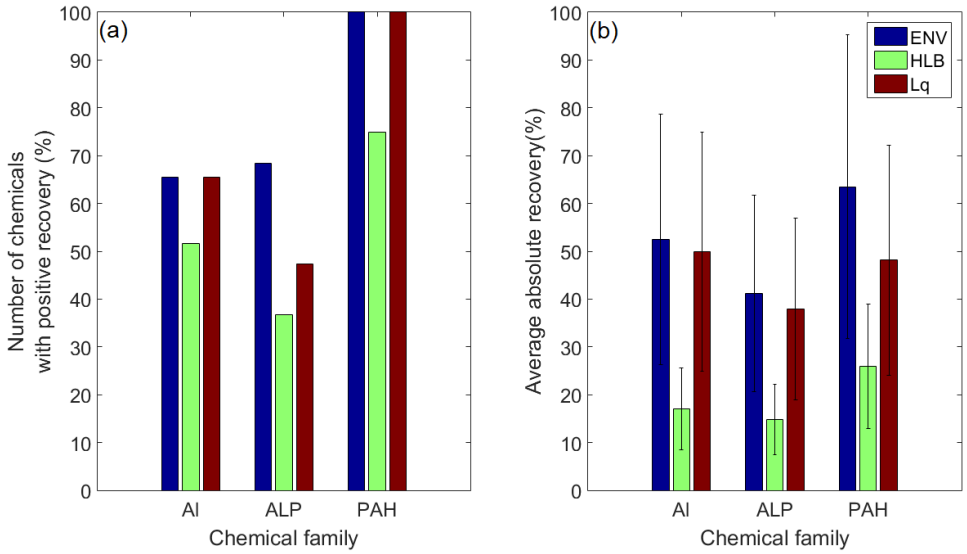


Figure 2: (a) Percentage of the target analytes with positive recoveries and (b) average absolute recoveries of target analytes with positive absolute recoveries. In panel "b" the error bars represent $\pm$ 2 $\times$ standard deviation of the recoveries for a chemical family via an extraction method.

Non-targeted Recovery Assessment

The F-ratio approach was employed for capturing the statistically meaningful features in the chromatograms. The features/fragments and/or molecular ions in the mass spectra that were causing the differentiation among investigated extraction methods were singled out through the combination of F-ratio analysis and apex detection. For the purpose of this proof of concept and to minimize false positives detection, we utilized a false positive detection probability value of 0.01% for the F-ratio, which corresponded to an F-ratio value of 3180, (Figure S6). Further optimization of the F-ratio value will be subject of future studies. This F-ratio value reduced the number of variables in the dataset by a factor of 95% and enabled us to focus only on the statistically significant features (Figure S7). After F-ratio correction, each chromatogram contained $\sim$ 2000 features. These features were a combination of redundant analytical signal (i.e. multiple features representing one unique feature, Figure S8), unresolved signal (i.e. signal which goes across a large section of chromatogram and does not

have a peak shape, Figure S7), and finally the noise, Figure S8. Those statistically significant features then were grouped, noise removed and unique features obtained by employing the apex detection algorithm. The apex detection resulted in 26 features which appeared to be highly relevant in differentiating the three extraction methods from each other. From those 2000 initial features, 67.4% were removed during the grouping process (i.e. redundant analytical signal), 28.9% of those features were unresolved signal and finally 3.7% of those features were classified as noise. The number of features belonging to redundant signals was in agreement with our expectations considering the sampling rate provided by the GC-Orbi (i.e. ~ 10 Hz based on the number of scans in an average peak). For example for each unique feature, on average, around 55 redundant analytical signals were observed that after grouping were represented by one unique feature (Figure S8). The unresolved features/signals and noise were excluded from the final unique feature list for further evaluation due to the difficulties in associating a chemical formula to them. Thus we used the relative recoveries (Eq. 1) of the final 26 unique features generated via the combination of F-ratio method and the apex detection algorithm for recovery assessment of different extraction methods.

The ENV method produced a relative recovery of 100% for all 26 unique features (i.e. the maximum averaged signal for all 26 unique features) whereas the Lq and HLB methods produced relative recoveries larger than zero for only 3 out of 26 unique features (Figure 3). The signal of 23 out of 26 unique feature was zero in the extraction methods Lq and HLB whereas a meaningful signal was produced in the chromatogram obtained from the ENV method (Figure S9). The low variability ($\leq 20\%$) observed for all the extraction methods and all the unique features further indicated the meaningfulness of these features. We also predicted the chemical formula of each of these unique features using the ChemCal online tool.[35] Additionally, another online tool (i.e. Isotope Distribution Calculator and Mass Spec Plotter[36]) was used to calculate the isotopic distribution of the predicted formula in order to provide further confirmation (Table S2). Based on the predicted chemical formulas of

the unique features (molecular fragment ions), most of those features contained one or more heteroatom (i.e. O, N, and S), which could be considered as an indication that these analytes were among the more "polar" compounds. Furthermore, the three features where the methods Lq and HLB produced larger than zero relative recoveries all appeared to be simple hydrocarbons without any heteroatoms. Therefore, the ENV method appeared to be more successful in extracting more "polar" components of PW. Further investigation is necessary in order to identify confidently the compounds which produced these unique features. None the less, the suggested approach was shown to be effective in capturing the relevant features that were causing the differentiation among the studied extraction methods. Also our results indicate the overall better performance of the ENV method in extracting PW compared to the other two methods. Finally, it should be noted that these 26 unique features are only a sub-sample of the unique statistically significant features in this dataset. In order to make sure that all the statistically significant features in differentiating these samples are captured an optimization of the F-ratio threshold is necessary. The optimization of this parameter will be subject of future studies.

The non-targeted approach was able to comprehensively evaluate the extraction recovery of PW via the three different methods. This method was effective where the traditional approaches (e.g. targeted method) failed to distinguish the best extraction method (e.g. the ENV and Lq methods were statistically similar).

Implications and Limitations

The combination of the F-ratio method and the apex detection algorithm was shown to be effective in isolating those features which allowed the differentiation of complex samples. In this study, we used this approach to evaluate the recovery of three widely used extraction methods for analysis of produced water. Our results suggested that one of the methods

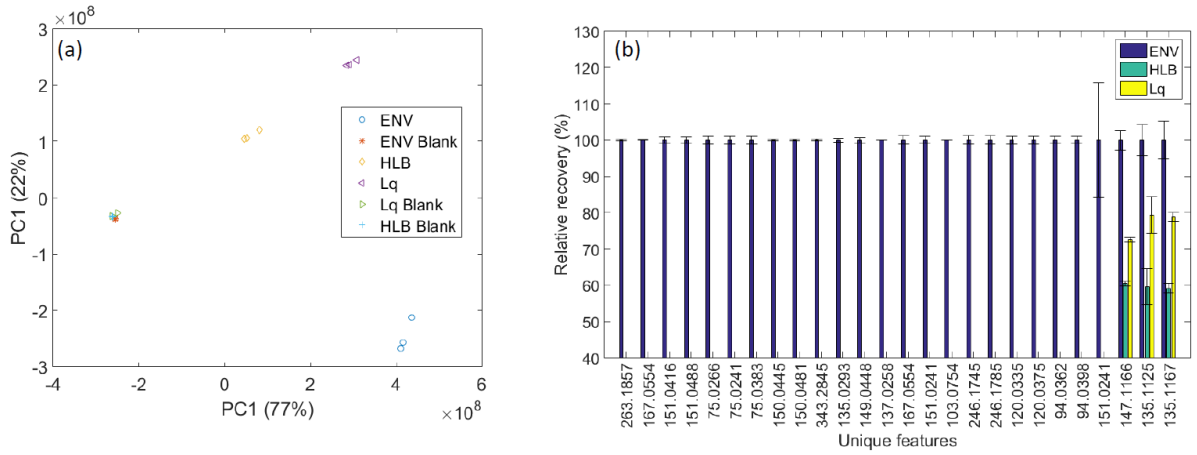


Figure 3: Depicting (a) the score plot of the first two principal components with percentage variability described and (b) relative recoveries of all 26 unique features using Eq. 1. The error bars in this figure represent ± standard deviation of the recoveries for a unique feature via an extraction method.

(i.e. using ENV method) performed far better than the other two methods, even though the traditional targeted approach failed to reveal the differences between these methods (i.e. ENV and Lq methods). This method captured the features that were statistically meaningful and also were extracted only using the ENV extraction method. Better understanding of the chemical space explored via each extraction method is highly relevant for the toxicity risk assessment, chemical processes/process engineering, and retrospective suspect and non-target screening. This method should enable analysts to evaluate qualitatively the extraction recovery of different methods and at the same time to explore the chemical space sampled via each extraction method. This would result in an optimized method, which would cover a wide area of chemical space. Additionally, the method proposed here has the potential to be applied to all cases where a change in the process may cause the generation of different outputs. For example, this method could be applied to the output of treated wastewater with different advanced oxidation processes, given the differences in the reaction pathways.

The main limitations of the present approach are the sensitivity towards high levels of variability, the computational cost, and the necessary MS resolution. For example, we cal-

culated the F-ratio values for the 65 target analytes in this study and those values ranged between 18 to 543, which were too small for them to be captured by the non-targeted approach. This was mainly caused by the high level of variability observed in the Lq extraction method (i.e. 198%). Therefore, this data processing method should be combined with the conventional targeted method in order to be able to evaluate its effectiveness, specially when expecting a larger level of variability in the dataset. In terms of the computational cost, the cloud computation (i.e. the use of a cluster of computers) should be considered in order to make these types of analysis possible in a timely fashion. The F-ratio method can be applied to data produced via both unit resolution MS[32,33] as well as high resolution data.[17] The necessary MS resolution for F-ratio analysis dependents on the level of complexity of the evaluated sample. In other words for highly complex samples such as produced water the F-ratio applied to low resolution GC-MS or LC-MS (i.e. unit mass) data may fail. Therefore, the analyst must choose the adequate MS resolution for the F-ratio analysis, based on the prior knowledge of the sample complexity. However, all considered, this approach (i.e. the combination of F-ratio method and the apex detection algorithm) appears to be a powerful tool for dealing with complex samples and chemical space problems.

Acknowledgement

The authors are thankful to the Research Council of Norway for the financial support of this project (RESOLVE, 243720). We are also grateful to Sintef, Trondheim and StatOil for providing us with the produced water samples.

Supporting Information Available

The Supporting Information including details regarding the sample preparation, analysis, steps taken during the data processing, and figures is available free of charge on the ACS

Publications website. Table S1 (an external file) containing the list of target analytes is also available free of charge on the ACS Publications website.

Associated Content

Author Information

Corresponding Author:

Saer Samanipour

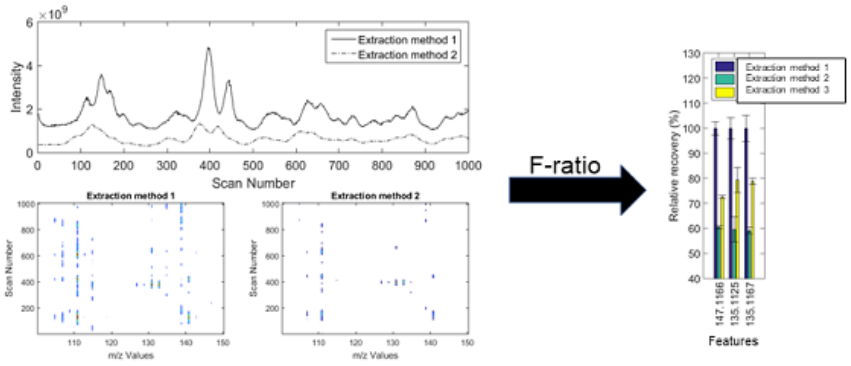
E-mail: saer.samanipour@niva.no

Phone: +47 98 222 087

Address: Norwegian Institute for Water Research (NIVA)

0349 Oslo, Norway

TOC Art



TOC Art

References

(1) Oetjen, K.; Giddings, C. G.; McLaughlin, M.; Nell, M.; Blotevogel, J.; Helbling, D. E.; Mueller, D.; Higgins, C. P. *Trends Environ. Anal. Chem.* **2017**, *15*, 12–23.

(2) Robson, W. J.; Sutton, P. A.; McCormack, P.; Chilcott, N. P.; Rowland, S. J. *Anal. Chem.* **2017**, *89*, 2919–2927.

(3) Jones, D.; Scarlett, A.; West, C.; Frank, R.; Gieleciak, R.; Hager, D.; Pureveen, J.; Tegelaar, E.; Rowland, S. *Chemosphere* **2013**, *93*, 1655–1664.

(4) Thomas, K. V.; Balaam, J.; Hurst, M. R.; Thain, J. E. *J. Environ. Monit.* **2004**, *6*, 593–598.

(5) Oil, N.; Gas, Environmental Report 2016. https://www.norskoljeoggass.no/no/Publikasjoner/MIljora
2016/, 2016.

(6) Thomas, K.; Langford, K.; Petersen, K.; Smith, A.; Tollefsen, K. *Environ. Sci. Technol.* **2009**, *43*, 8066–8071.

(7) Thomas, K. V.; Balaam, J.; Hurst, M. R.; Thain, J. E. *Environ Toxicol. Chem.* **2004**, *23*, 1156–1163.

(8) Balaam, J. L.; Chan-Man, Y.; Roberts, P. H.; Thomas, K. V. *Environ. Toxicol. Chem.* **2009**, *28*, 1159–1167.

(9) Booth, A. M.; Scarlett, A. G.; Lewis, C. A.; Belt, S. T.; Rowland, S. J. *Environ. Sci. Technol* **2008**, *42*, 8122–8126.

(10) Booth, A. M.; Sutton, P. A.; Lewis, C. A.; Lewis, A. C.; Scarlett, A.; Chau, W.; Widdows, J.; Rowland, S. J. *Environ. Sci. Technol.* **2007**, *41*, 457–464.

(11) Melbye, A. G.; Brakstad, O. G.; Hokstad, J. N.; Gregersen, I. K.; Hansen, B. H.; Booth, A. M.; Rowland, S. J.; Tollefsen, K. E. *Environ. Toxicol. Chem.* **2009**, *28*, 1815–1824.

(12) Spanik, I.; Machynakova, A. *J. Sep. Sci.* **2018**, *41*, 163–179.

(13) Luek, J. L.; Gonsior, M. *Water research* **2017**, *123*, 536–548.

(14) Staš, M.; Chudoba, J. *Chemické listy* **2017**, *111*, 628–636.

(15) Headley, J. V.; Peru, K. M.; Barrow, M. P. *Mass spectrometry reviews* **2016**, *35*, 311–328.

(16) Radovic, J. R.; Thomas, K. V.; Parastar, H.; Díez, S.; Tauler, R.; Bayona, J. M. *Environmen. Sci. Technol.* **2014**, *48*, 3074–3083.

(17) Samanipour, S.; Reid, M. J.; Thomas, K. V. *Anal. Chem.* **2017**, *89 (10)*, 5585–5591.

(18) Schollee, J. E.; Schymanski, E. L.; Avak, S. E.; Loos, M.; Hollender, J. *Anal. Chem.* **2015**, *87*, 12121–12129.

(19) Brereton, R. G. *Applied chemometrics for scientists*; John Wiley & Sons, 2007.

(20) Parsons, B. A.; Marney, L. C.; Siegler, W. C.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2015**, *87*, 3812–3819.

(21) Baz-Lomba, J. A.; Reid, M. J.; Thomas, K. V. *Anal. Chem. acta* **2016**, *914*, 81–90.

(22) Samanipour, S.; Baz-Lomba, J. A.; Alygizakis, N. A.; Reid, M. J.; Thomaidis, N. S.; Thomas, K. V. *J. Chromatogr. A* **2017**, *1501 (2017)*, 68–78.

(23) Baker, D. R.; Kasprzyk-Hordern, B. *Journal of Chromatography A* **2011**, *1218*, 8036–8059.

(24) Fatta, D.; Achilleos, A.; Nikolaou, A.; Meric, S. *TrAC Trends in Analytical Chemistry* **2007**, *26*, 515–533.

(25) Noro, Norwegian Oil and Gas recommended guidelines for sampling and analysis of produced water, translated version. 2003.

(26) Thomas, K. V.; Langford, K.; Petersen, K.; Smith, A. J.; Tollefsen, K. E. *Environ. Sci. Technol.* **2009**, *43*, 8066–8071.

(27) Statoil, N. Heidrun oil platform. https://www.statoil.com/en/what-we-do/norwegian-continental-shelf-platforms/heidrun.html, 2017.

(28) Samanipour, S.; Langford, K.; Reid, M. J.; Thomas, K. V. *J. Chromatogra. A* **2016**, *1463*, 153–161.

(29) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534–2536.

(30) MATLAB version 9.1 Natick, Massachusetts: The MathWorks Inc.,

(31) Vinaixa, M.; Samino, S.; Saez, I.; Duran, J.; Guinovart, J. J.; Yanes, O. *Metabolites* **2012**, *2*, 775–795.

(32) Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. *Anal. Chem.* **2006**, *78*, 5068–5075.

(33) Pierce, K. M.; Hope, J. L.; Johnson, K. J.; Wright, B. W.; Synovec, R. E. *J. Chromatogr. A* **2005**, *1096*, 101–110.

(34) Breslow, N. *Biometrika* **1970**, *57*, 579–594.

(35) Patiny, L.; Borel, A. ChemCalc: a building block for tomorrow's chemical infrastructure. 2013.

(36) (SIS), S. I. S. Isotope Distribution Calculator and Mass Spec Plotter. http://www.sisweb.com/mstools/isotope.htm, 2015; Online tool.