Non-target screening of organic compounds in offshore produced water by GC×GC-MS

Sofie Nitsche Bergfors¹, Khoa Huynh¹, Annette Eva Jensen¹, Jonas Sundberg¹

¹ Danish Hydrocarbon Research and Technology Centre, Technical University of Denmark, Kgs. Lyngby, Denmark

Corresponding Author:

Jonas Sundberg¹

Email address: jonsun@dtu.dk

ORCID 0000-0002-6308-5270

Abstract

Produced water is the largest by-product of oil and gas production. At off-shore installations, the produced water is typically reinjected or discharged into the sea. The water contains a complex mixture of dispersed and dissolved oil, solids and inorganic ions. A better understanding of its composition is fundamental to 1) improve environmental impact assessment tools and 2) develop more efficient water treatment technologies. The objective of the study was to screen produced water sampled from a producing field in the Danish region of the North Sea to identify any containing organic compounds. The samples were taken at a test separator and represent an unfiltered picture of the composition before cleaning procedures. The analytes were isolated by liquid-liquid extraction and derivatized using a silylation reagent to increase the volatility of oxygenated compounds. The final extracts were analyzed by comprehensive multi-dimensional gas chromatography coupled to a high-resolution mass spectrometer. A non-target processing workflow was implemented to extract features and quantify the confidence of library matches by correlation to retention indices and the presence of molecular ions. Approximately 120 unique compounds were identified across nine samples. Of those, 15 were present in all samples. The main types of compounds are aliphatic and aromatic carboxylic acids with a small fraction of hydrocarbons. The findings have implications for developing improved environmental impact assessment tools and water remediation technologies.

Introduction

Produced water is by volume the largest byproduct in oil and gas production. For reservoirs on the Danish continental shelf, where production is often supported by water flooding, the volume of produced water typically exceeds the volume of oil (The Danish Energy Agency, 2014). This water contains a complex mixture of inorganic and organic compounds. Implementation of offshore water management strategies is challenging due to the large volumes and practical limitations on infrastructure. Current procedures consist of cleaning of the produced water, followed by discharge into the sea or reinjection in producing or disposal wells (Lyngbaek & Blidegn, 1991; Røe & Johnsen, 1996; Røe Utvik, 1999; Durell et al., 2004). For discharge, the OSPAR Convention, which was set in to force in 1998, sets a limit of 30 mg of dispersed oil per liter of water as an annual average (OSPAR Commission, 2001). There is currently no limit on the dissolved compounds. Furthermore, detailed knowledge on their structure and abundance is lacking.

In simplified terms, produced water can be seen as the product of aqueous extraction of crude oil by mixing within the reservoir and during production. In reality, the process is complex and the composition is highly dependent on reservoir properties, field history and water injection strategies (Bergfors, Schovsbo & Feilberg, 2020). For example, studies have shown that salinity influences to the organic content, likely caused by a salting-out effect (Barth, Borgund & Riis, 1990; Barth, 1991; Barth & Riis, 1992; Endo, Pfennigsdorff & Goss, 2012; Dudek et al., 2020). At the producing platform, a gravitational separation of oil, water and gas is carried out. The resulting water phase contains dissolved and dispersed oil droplets. The latter is largely removed by physical methods, i.e. separators, hydrocyclones and gas flotation tanks/degassers implemented in series on the platform, typically resulting in levels below 10 mg/L at the discharge point (Meldrum, 1988; Saththasivam, Loganathan & Sarp, 2016; Durdevic & Yang, 2018). In contrast, the removal of dissolved organics requires chemical treatment, e.g. degradation via advanced oxidation processes (AOPs) (Jiménez et al., 2019; Lin et al., 2020; Liu et al., 2021). This is challenging to implement on offshore installations due to safety and the requirement of low residency times. Furthermore, the current generation of AOPs are mainly based on Fenton's reagent which produces toxic chlorinated byproducts when applied to saline water (Kiwi, Lopez & Nadtochenko, 2000; De Laat & Le, 2006; Sirtori et al., 2012). The goal of produced water management is not necessarily zero discharge but zero *harmful* discharge. Improved knowledge of its composition is thus not only beneficial to develop environmental impact assessment tools. The information may also be used to develop improved and targeted treatment methods, i.e specific removal of harmful components, and not total organic content. This has the potential to increase efficiency and reduce the cost of water management strategies.

Ultimately, the toxicity of the produced water is a complex function based on the structure and concentration of both the dispersed and dissolved oil components (Strømgren et al., 1995; Bakke, Klungsøyr & Sanni, 2013; Niu et al., 2016; Lofthus et al., 2018). Furthermore, inorganic species such as heavy metals should be taken into consideration. We argue that to fully understand the effect of organic compounds in produced water discharge, we must 1) carry out full structural characterization studies and 2) differentiate between dispersed and dissolved species and 3) develop methods to integrate the data into environmental impact assessment tools. Although numerous compounds in produced water have been identified as part of ongoing water quality studies, a large part of the chemical space remains unknown. Recent advances in analytical technology (i.e. multi-dimensional chromatography, high-resolution mass spectrometry) have allowed more in-depth studies of the organic composition (Samanipour et al., 2019, 2020; Sørensen et al., 2019; Dudek et al., 2020). However, these studies mainly focus on general classes of naphthenic acids and are based on formula calculations of exact masses without further structural elucidations. Sørensen et. al. described a non-target screening study but did not provide details beyond compound classes (i.e. phenols vs. naphthalenes). Thus, a large knowledge gap in the identity of dispersed and dissolved organics still exists.

To narrow this gap we have carried out a non-target screening study. The objective was not to evaluate the efficiency of implemented cleaning procedures but to obtain an “unfiltered” picture of the water composition immediately after production. Several samples were obtained from the test separator of a producing platform in the Danish North Sea. The organics were extracted by liquid-liquid (LLE) extraction using dichloromethane (DCM). The extracts were derivatized to increase the volatility of oxygenated compounds and analyzed using comprehensive 2D-gas chromatography (GC×GC) coupled with a high-resolution quadrupole-time of flight mass spectrometer (MS). One advantage of using GC-MS as compared to liquid-chromatography (LC)-MS is the possibility of identification by spectral matching using commercially available libraries. In contrast, LC-MS is heavily dependent on user-generated libraries based on reference compounds as the ionization efficiency is much more dependent on experimental conditions (Schymanski et al., 2015). Secondly, by using multidimensional GC we were able to resolve more compounds than would be possible with conventional LC. Approximately 1500 compounds were detected per sample, with 120 unique compounds tentatively identified across all samples. To increase the confidence of library matches, a data processing workflow that implemented retention indices and accurate mass matching was implemented. The data were analyzed to obtain a broad knowledge of the types of compounds present, their relative abundance as well as common compounds present across all samples.

Materials & Methods

Chemicals and reagents

Benzoic acid, phenol, cyclohexanecarboxylic acid, octanoic acid, dichloromethane (LiChroSolv, Merck), *n*-hexane (SupraSolv for gas chromatography MS, Merck), magnesium sulfate (ReagentPlus, Redi-Dri, Sigma Aldrich), N,O-bis(trimethylsilyl)trifluoroacetamide containing 1% of trimethylchlorosilane (BSTFA+TMCS, Supelco) were used as received. Deuterated internal standards (naphthalene- d_8 , acenaphthene- d_{10} , phenanthrene- d_{10} , chrysene- d_{12} , Supelco analytical standards) were used to monitor retention time shifts. 1D retention index calibration was performed using a linear C7 to C30 saturated alkanes mixture (Supelco, TraceCERT, Sigma Aldrich). High purity water was obtained from a Milli-Q Advantage A10 unit. All chemicals and reagents were used as received.

Sampling and sample preparation

Nine produced water samples were acquired from a producing field (water-injected) located in the Danish region of the North Sea. The sampling campaign took place between June 2018 and February 2019. Water samples were collected at the test separator on the production platform following protocols established by the operator. No cleaning or further processing of the samples was carried out at the platforms. The samples were received in plastic bottles, aliquoted, and immediately treated with dilute hydrochloric acid (18%, 1 mL per 100 mL sample) for a final pH < 2. The samples were stored in the dark at 4 °C until extraction and analyses.

All samples were extracted in three experimental replicates using separate glassware. Samples were filtered through a 0.45 μm PTFE-filter to remove solids and particles. 50 mL of produced water was extracted with DCM (50 mL). The organic phase was washed with saturated aqueous sodium chloride (50 mL) and carefully removed in vacuo. The residue was reconstituted in *n*-hexane (1.5 mL) and dried over $MgSO_4$. An aliquot (1000 μL) of the sample was transferred to a 2 mL vial, combined with deuterated internal standards (for monitoring of retention time stability), combined with BSTFA+TMCS (50 μL) and incubated at 70 °C for 30 minutes, whereafter, it was allowed to return to room temperature. The derivatized sample was further diluted ten-fold with *n*-hexane and immediately analyzed on GC $\times$ GC-MS.

Extraction recovery and reproducibility

A synthetic produced water containing five representative model compounds (benzoic acid, phenol, 2-naphthoic acid, cyclohexanecarboxylic acid, and octanoic acid, each at 5 ppm, total organics 25 ppm) in simulated formation water (see supplemental information) was prepared to establish variability in the sample preparation protocol and instrumental analysis. The synthetic water was extracted four times in

two batches. Three procedural blanks were prepared to establish background levels and experimental sources of contamination.

GC×GC-MS

GC×GC-MS data were acquired using an Agilent 7890B GC coupled to a 7200B QTOF high-resolution mass spectrometer (Agilent Technologies, Palo Alto, CA, USA). The system was equipped with a Zoex ZX-2 thermal modulator (Zoex Corporation, Houston, TX, USA). The separation was achieved using a combination of an Agilent DB-5MS UI (^1D , 30 m, 0.25 mm i.d., 0.25 μm d_f) and a Restek Rxi-17Sil MS (^2D , 2 m, 0.18 mm i.d., 0.18 μm d_f) capillary columns connected using a SilTite μ -union. The oven was temperature programmed as follows; 1 min hold-time at 50 °C, ramp to 320 °C, 3 °C min^{-1} in constant flow mode (1 mL/min). The modulation period was set to 3 s with a 400 ms hot-jet duration. The MS transfer line was held at 280 °C. The MS acquired spectra in electron ionization mode (70 eV) with a mass range of 45 – 500 and an acquisition speed of 50 Hz. The instrument was operated in its 2 GHz sampling rate mode to increase the dynamic range. Automatic mass calibration was performed for every 5th sample (approximately 7.5 hours).

Data processing

Peak detection and library search were performed using GC Image v2.8.3 (Zoex, Houston, TX). Mass spectra were matched against the NIST Mass Spectral Library (National Institute of Standards and Technology, 2017 edition) with a minimum match factor of 700. All compound tables were exported as comma-separated texts for external processing. A data processing workflow was implemented in Python (Python Software Foundation. Python Language Reference, version 3.7.4. Available at <http://www.python.org>). The script is available as supplementary information under data availability.

Figure 1. Photo of four samples as received. Only clear samples were used in the study, while samples containing oil were excluded as they contained large amounts of dispersed oil droplets in the aqueous phase (see Figure 2).

Results

Sample preparation and analyses

Previous studies on dissolved organics in oilfield produced water have employed LLE or solid-phase extraction (SPE) (Thomas et al., 2009; Kovalchik et al., 2017; Barros et al., 2018; Samanipour et al., 2019). For produced water, LLE using DCM has been shown to have a similar recovery as SPE (Samanipour et al., 2019). The main difference was observed in the size distribution, where larger species tend to have a higher recovery using SPE due to their low solubility in DCM. LLE is ‘non-discriminative’ in comparison to SPE which is used to fractionate compound classes based on adsorption characteristics. Thus, it allowed us to extract the broad range of organics present in produced water. Furthermore, the Norwegian Oil and Gas specialist network recommends LLE for the quantification of phenols in produced water (Norwegian Oil and Gas, 2012). Thus, it would allow us to ‘see’ what is missing using routine targeted analyses. The water samples had varying levels of oil, likely due to the sampling and status of the test separator. Samples that contained a clear separate layer of oil, including smaller amounts, were excluded from the study to minimize the risk of contamination (Figure 1). Optical inspection of these samples showed that they contained large amounts of dispersed oil droplets, even when sampling below the oil layer (Figure 2). All samples were filtered through a 0.45 μm PTFE-filter to remove insoluble material and particles. A small aliquot (50 mL) of each sample was extracted in triplicate using an equivalent volume of DCM.

Figure 2. Optical microscope image of produced water samples free of (left) and containing dispersed microdroplets (right).

The dissolved components of produced water were largely expected to be oxygenated organics, i.e. alcohols (mainly phenols) and acids. To increase the volatility of the aforementioned compounds, the samples were silylated. After removal of the solvent in-vacuo, each sample was reconstituted in 1.5 mL of *n*-hexane, treated with BSTFA-TMCS and incubated at 70 °C for 30 minutes. After derivatization, the samples were further diluted ten-fold, meaning that the samples ultimately were concentrated approximately three-fold (from 50 mL to 15 mL). The final concentration factor was chosen as a compromise between the detection of trace compounds and avoiding column and detector overload. Due to the large concentration range of our analytes, it would be beneficial to run each sample at multiple dilution factors. However, due to the long run time (90 minutes) and the number of extraction replicates (3) this was not feasible due to time constraints. It is important to remember that this strategy has two effects; 1) high concentration compounds may overload the detector with spectral skewing and poor library match as result, and 2) trace-compounds may be diluted to a level below the limit of detection.

Model compound	Recovery (%)	%RSD, model water	%RSD, standard	%RSD, real-life sample*
Phenol	89.6	23.1	8.9	23.3
Cyclohexanecarboxylic acid	63.8	13.1	5.7	1.8
Benzoic acid	35.9	24.7	1.9	11.3
Octanoic acid	55.0	9.6	4.4	4.2
2-Naphthoic acid	88.4	9.8	3.0	5.2
Internal standard	Avg. RT I (min.)	Avg. RT II (sec.)	RT I %RSD	RT II %RSD
Naphthalene-D8	19.399	1.860	0.0004	0.8576
Acenaphthene-D8	32.199	2.236	0.0002	0.6218
Phenanthrene	43.399	2.671	0.0002	0.5178

Table 1. Recovery and RSD values. Recovery values are calculated as $(\text{peak volume}_{\text{model water}}/\text{peak volume}_{\text{standard}})$ * 100 (n = 4 for model water and standard, 3 for real-life sample). Retention time deviations were calculated based on

the retention of internal standards across all samples ($n = 27$). * Peak volume values were extracted from a representative sample where the compound had been identified.

Three procedural blanks were extracted and analyzed to determine background levels and potential sources of contamination. Minor amounts of fatty acids were detected in addition to a series of polysiloxanes (discrete peaks, not common column bleed). The source of the latter could not be identified but as it lacked retention in the 2D it did not interfere with our analytes of interest. To establish recovery and repeatability, a model produced water was prepared by spiking four organic acids and phenol into a synthetic produced water. All model compounds were detected in one or more real-life water samples. The model water was extracted in four experimental replicates. The recovery values were calculated based on peak volumes in comparison to those obtained by analyzing pure stock solutions (Table 1). The recovery varied from 36% for benzoic acid up to 90% for phenol. Considering the multi-step sample extraction, including a derivatization reaction, this was deemed acceptable. The relative-standard deviation of the model compounds as detected in a representative real-life sample varied from 2% to 23% (calculated from three experimental replicates).

Chromatography

Crude oil is an ultracomplex mixture of saturated and aromatic hydrocarbons with a smaller fraction of N,S,O-containing compounds (Marshall & Rodgers, 2004; Hsu et al., 2011; Palacio Lozano et al., 2020). This complexity will be reduced but reflected in the produced water. Based on previous studies of Danish oils we know that the dominant oxygenated species belong to the O1 and O2 classes with a large diversity in aromaticity (Sundberg & Feilberg, 2020). Thus, the compositional variation in the produced water was assumed to be dominated by carbon number and level of saturation. Ultimately, the boiling point range was assumed to be larger than the variation in saturated versus aromatic structures. Therefore, we choose to use a non-polar column in the 1D (providing the highest separation power) with a shorter medium polarity column in the 2D . Conventional polar columns are based on polyethyleneglycol (PEG) chemistry and are incompatible with silylation reagents. Therefore, we choose to use a 50%-phenyl-type column where aromaticity is the largest factor affecting retention. By using this column combination, the retention in both dimensions will increase with aromaticity. For example, cyclohexane acetic acid has a retention of 17.2 min/0.84 s in $^1D/^2D$ as compared to 24.6 min/1.33 s for benzeneacetic acid (measured as the corresponding trimethylsilyl esters). In contrast, alkylation of a core aromatic or saturated structure will only affect retention in the 1D . For example, phenol and its alkylated homologs (methyl, ethyl and propyl) have 1D retention times of 13.4, 16.9, 19.9 and 23.3 minutes, respectively, where the retention in 2D is within 0.89 to 0.95 seconds. By investigation of a typical chromatogram, two things become obvious; 1)

the desired separation of saturates, mono- and diaromatics is achieved and 2) a large portion of the 2D space is relatively uncopied due to the small number of polycyclic species where the majority of aromatics are benzene derivatives.

A $3\text{ }^{\circ}\text{C min}^{-1}$ temperature gradient was found to be the optimal compromise between peak width, resolution and run time. At this rate, the typical peak width was 10 s in the first dimension. Thus, using a modulation period of 3 s allowed us to obtain the recommended minimum of three modulations per peak (Murphy, Schure & Foley, 1998). Due to instrumental complexity and long run times, retention time shifts are commonly observed both in inter-sample and inter-batch runs. Deuterated PAH standards were used to monitor retention times over time. The $^1D/^2D$ retention time variability is presented in Table 1. Both 1D and 2D were shown to be stable over the whole analyses run, covering 27 seven injections (9 samples with three experimental replicates) and 7 days.

Figure 3. Typical 2D chromatogram showing the distribution of peaks with a heavy saturation of compounds between 0.5 and 1.5 s in the 2D .

Non-target screening and compound identification

Approximately 1500 compounds were detected in each sample. A data processing workflow was implemented to sort, organize and score the data. A schematic representation is presented in Figure 3. The associated files are available as supplementary information. The detected features were matched against the NIST EI Mass Spectral Library (2017). All compounds with a match factor below 700 were removed.

To increase the annotation confidence, two additional factors were included; retention index (Kovats) and the presence of the molecular ion. To quantify the identification confidence the following scoring rules were implemented:

1. A match factor above 800 gives a score of 10 points.
2. A match factor between 700 and 800 gives a score of 5 points.
3. A retention index match within 50 units gives a score of 5 points.
4. The detection of the molecular ion within 20 ppm mass accuracy gives a score of 5 points.

Figure 4. Schematic of the data processing workflow.

A total score of above 10 was (i.e. match factor of a minimum of 700 and either a retention index and or molecular ion match) was classified as a probable match, whereas a total score below 10 was classified as a tentatively identified structure (Schymanski et al., 2014). To reduce experimental errors, only features that were present in all three replicates were included in the final table of compounds. Furthermore, all duplicate compounds were removed. This is a crude step that inherently removes, for example, isomeric species which would not be differentiated using automatic library search. However, for our purpose the identification of isomers is not an inherent goal. The aim of the study was to identify the presence of broad compound-class types and dominant species. The correct identification of isomeric species or species with highly similar fragmentation patterns requires manual intervention and is beyond the scope of this study.

The workflow reduced the number of features from approximately **1)** 1500 detected to **2)** 200 library hits with match factor > 700, to **3)** 50 - 70 when with duplicates based on name were removed. Ultimately, the merging of inter-sample features resulted in 120 unique hits across all samples. Of those, 42 had the maximum score of 20 (i.e. match factor > 800, molecular ion detected and retention index within 50 units). Near all (87%) identified compounds are oxygen-containing, with amines, sulfides and hydrocarbons being the remaining constituents. Only 15 compounds (after removal of internal standards and background species) were present in all samples (Table 2).

Discussion

A few papers have previously described non-target screening of produced water, primarily from the Norwegian continental shelf. Sørensen et. al. described the characterization of unpurified DCM extracts with comparisons to the polar and non-polar fractions as isolated by SPE (Sørensen et al., 2019). In their study, a high concentration of hydrocarbons including naphthalenes and linear alkanes was detected in the produced water. This is in contrast to our study, where only a few hydrocarbons, in relative trace amounts, were observed as measured by extracted ion chromatograms of known species. The aqueous solubility of saturated hydrocarbons is low. However, hydrocarbons could be present in the water phase as dispersed oil droplets. We observed that the presence of such droplets in a water sample is heavily dependent on the oil-water ratio, and that optical inspection using a microscope was required for their detection. Oil droplets were not observed in the samples included in our study. Thus, we believe that the majority of detected compounds can be classified as dissolved organics.

Only a small fraction of the detected compounds were identified with an acceptable confidence level. 36 unique compounds (across all samples) received the maximum identification score, i.e. a match factor > 800, retention index match and detection of the molecular ion. The match factor had the most severe impact on feature reduction. Lowering the match factor limit to >600 increased the number of tentatively identified features by approximately 40% (compared to match factor > 700). A manual evaluation showed that although several hits were chemically reasonable based on structure and retention index, the lower limit also led to multiple apparent false positives. Added confidence to questionable identities could be obtained by a corroborative study using soft ionization techniques, i.e. chemical ionization or low voltage electron ionization, where the molecular ion is better observed for non-aromatic species.

Figure 5. The molecular structures of six representative compounds from the subset that were detected in a minimum of 6 out of 9 samples. LogP/XLogP values were obtained from the PubChem database. The reported retention times are of the corresponding trimethylsilyl derivatives.

Looking at the obtained data, two conclusions were made; 1) samples were dominated by oxygenated organics, and 2) sample-to-sample variation was high, both in terms of composition and relative concentrations. Oxygenated hydrocarbons form during diagenesis but may also be the result of microbial and/or chemical processes during oil production (Aitken, Jones & Larter, 2004; Head, Gray & Larter, 2014; Pannekens et al., 2019). The oxygenation leads to high partitioning into the aqueous phase during oil-water separation, and these compounds will likely require attention when developing successful water management technologies. Approximately 50% of the compounds were aromatic, primarily benzene-derivatives with few naphthalenes present. Some of the identified compounds are suspected residual production chemicals, e.g. hexadecanol which does not occur naturally in crude oil. Only 15 compounds were identified across all 9 samples (Table 3). Even when comparing two samples that were sampled from the same well and only three days apart, the difference was substantial. The intra-sample variation is likely an effect of sampling and test separator status. This leads to a differing oil-to-water ratio, and thus the presence of dispersed and dissolved oil components

Compound	MS-ready name	Formula	Mol. Weight	XlogP	SMILES	Classification	#Samples
(±)-2-Phenylpropanoic Acid, trimethylsilyl ester	2-phenylpropanoic acid	C9H10O2	150.17	1.9	CC(C1=CC=CC=C1)C(=O)O	Aromatic acid	9
1-Naphthoic acid, TMS derivative	naphthalene-1-carboxylic acid	C11H8O2	172.18	3.1	C1=CC=C2C(=C1)C=CC=C2C(=O)O	Aromatic acid	9
3,4-Dimethylbenzoic acid, TMS derivative	3,4-dimethylbenzoic acid	C9H10O2	150.17	2.7	CC1=C(C=C(C=C1)C(=O)O)C	Aromatic acid	9
Trimethylsilyl 2,3-dimethylbenzoate	2,3-dimethylbenzoic acid	C9H10O2	150.17	2.8	CC1=C(C(=CC=C1)C(=O)O)C	Aromatic acid	9
Trimethylsilyl 4-propylbenzoate	4-propylbenzoic acid	C10H12O2	164.2	3.4	CCCC1=CC=C(C=C1)C(=O)O	Aromatic acid	9
4-tert-Butylphenol, TMS derivative	4-tert-butylphenol	C10H14O	150.22	3.3	CC(C)(C)C1=CC=C(C=C1)O	Aromatic alcohol	9
3-Methyl-1-cyclohexanecarboxylic acid, trimethylsilyl ester (stereoisomer 2)	3-methylcyclohexane-1-carboxylic acid	C8H14O2	142.2	2.1	CC1CCCC(C1)C(=O)O	Saturated acid	9
3-Methylbutanoic acid, TMS derivative	3-methylbutanoic acid	C5H10O2	102.13	1.2	CC(C)CC(=O)O	Saturated acid	9
Cyclohexanecarboxylic acid, TMS derivative	2-cyclohexylacetic acid	C8H14O2	142.2	2.5	C1CCC(CC1)CC(=O)O	Saturated acid	9
Cyclohexanecarboxylic acid, TMS derivative	cyclohexanecarboxylic acid	C7H12O2	128.17	1.9	C1CCC(CC1)C(=O)O	Saturated acid	9
Cyclopentylcarboxylic acid, TMS derivative	cyclopentanecarboxylic acid	C6H10O2	114.14	1.3	C1CCC(C1)C(=O)O	Saturated acid	9
Heptanoic acid, TMS derivative	heptanoic acid	C7H14O2	130.18	2.5	CCCCCCC(=O)O	Saturated acid	9
Nonanoic acid, TMS derivative	nonanoic acid	C9H18O2	158.24	3.5	CCCCCCCCC(=O)O	Saturated acid	9
Octanoic acid, TMS derivative	octanoic acid	C8H16O2	144.21	3	CCCCCCCC(=O)O	Saturated acid	9
1-Hexadecanol, TMS derivative	hexadecan-1-ol	C16H34O	242.44	7.3	CCCCCCCCCCCCCCCCCO	Saturated alcohol	9
Benzenepropanoic acid, TMS derivative	3-phenylpropanoic acid	C9H10O2	150.17	1.8	C1=CC=C(C=C1)CCC(=O)O	Aromatic acid	8
m-Toluic acid, TMS derivative	3-methylbenzoic acid	C8H8O2	136.15	2.4	CC1=CC(=CC=C1)C(=O)O	Aromatic acid	8
2,4-Di-tert-butylphenol	2,4-ditert-butylphenol	C14H22O	206.32	4.9	CC(C)(C)C1=CC(=C(C=C1)O)C(C)(C)C	Aromatic alcohol	8
3-Ethylphenol, TMS derivative	3-ethylphenol	C8H10O	122.16	2.4	CCC1=CC(=CC=C1)O	Aromatic alcohol	8
4-Isopropylphenol, TMS derivative	4-propan-2-ylphenol	C9H12O	136.19	2.9	CC(C)C1=CC=C(C=C1)O	Aromatic alcohol	8
o-Cresol, TMS derivative	2-methylphenol	C7H8O	108.14	2	CC1=CC=CC=C1O	Aromatic alcohol	8
2-Methylbutanoic acid, TMS derivative	2-methylbutanoic acid	C5H10O2	102.13	1.2	CCC(C)C(=O)O	Saturated acid	8
3-Methylvaleric acid, TMS	3-methylpentanoic acid	C6H12O2	116.16	1.6	CCC(C)CC(=O)O	Saturated acid	8
2-Octanol, TMS derivative	octan-2-ol	C8H18O	130.23	2.9	CCCCCCC(C)O	Saturated alcohol	8
Benzenecarboxylic acid, TMS derivative	2-phenylacetic acid	C8H8O2	136.15	1.4	C1=CC=C(C=C1)CC(=O)O	Aromatic acid	7
Benzenebutanoic acid, TMS derivative	4-phenylbutanoic acid	C10H12O2	164.2	2.4	C1=CC=C(C=C1)CCCC(=O)O	Aromatic acid	7
m-Cresol, TMS derivative	3-methylphenol	C7H8O	108.14	2	CC1=CC(=CC=C1)O	Aromatic alcohol	7
2-Hydroxy-4-methylquinoline, trimethylsilyl ether	4-methyl-1H-quinolin-2-one	C10H9NO	159.18	1.2	CC1=CC(=O)NC2=CC=CC=C12	Aromatic amine/alcohol	7
4-Methylvaleric acid, TMS derivative	4-methylpentanoic acid	C6H12O2	116.16	1.4	CC(C)CCC(=O)O	Saturated acid	7
2-Ethylphenol, TMS derivative	2-ethylphenol	C8H10O	122.16	2.5	CCC1=CC=CC=C1O	Aromatic alcohol	6
4-Trimethylsilylphenol	phenol	C6H6O	94.11	1.5	C1=CC=C(C=C1)O	Aromatic alcohol	6
Benzoic acid, 4-ethoxy-, ethyl ester	ethyl 4-ethoxybenzoate	C11H14O3	194.23	3.2	CCOC1=CC=C(C=C1)C(=O)OCC	Aromatic ester	6
Naphthalene, 1,7-dimethyl-	1,7-dimethylnaphthalene	C12H12	156.22	4.4	CC1=CC2=C(C=CC=C2C=C1)C	Aromatic hydrocarbon	6
3-Methyl-1-cyclohexanecarboxylic acid, trimethylsilyl ester (stereoisomer 1)	3-methylcyclohexane-1-carboxylic acid	C8H14O2	142.2	2.1	CC1CCCC(C1)C(=O)O	Saturated acid	6
Pentanoic acid, TMS derivative	pentanoic acid	C5H10O2	102.13	1.4	CCCCC(=O)O	Saturated acid	6
2-(1-Adamantyl)ethanol, TMS derivative	2-(1-adamantyl)ethanol	C12H20O	180.29	3.4	C1C2CC3CC1CC(C2)(C3)CCO	Saturated alcohol	6

Table 2. A list of compounds that were detected in a minimum of 6 of 9 samples. * The MS-ready name corresponds to the non-derivatized parent compound.

Calculated XlogP values were obtained from the PubChem database. #Samples corresponds to the number of samples of which the compound was detected in.

Conclusions

The composition of produced water is highly complex with several unknowns. Our study aimed to narrow this gap by a broad identification of dissolved organics. The implemented identification workflow excluded approximately 95% of the detected compounds (1000 – 1500 per sample), resulting in 50 – 80 tentatively identified compounds per sample. This demonstrates both the power and pitfalls of non-target screening; more than 100 compounds were identified with an acceptable level of confidence, and more than 1000 compounds remain unknown. To our knowledge, this is the most comprehensive list of identified compounds in produced water that has been publicly published. However, being a screening study, quantification was not carried out for any compounds. As this is an important factor for environmental assessment, the obtained compound lists should be used to develop targeted methods to look at absolute concentrations. Furthermore, it would be beneficial to reduce the number of unknowns by using complementary techniques (e.g. HPLC-MS and other soft ionization methods), improved custom libraries and in-silico mass spectral prediction.

When performing environmental impact assessments, both structure and concentration have to be taken into consideration (Tang et al., 2019). A large part of the detected compounds are present at trace levels. Their concentration will be further reduced at discharge to sea where a rapid dilution to a large body of water occurs. However, potential cocktail effects where the combined effect of a series of micro-pollutants is harmful but not the single species should be accounted for (Di Poi et al., 2018). Here, it would be highly beneficial to link non-target screening studies with toxicological measurements. Ultimately, we hope that our study contributes a small piece of the puzzle, and a stepping stone towards further studies to uncover the full picture.

Data availability

Associated data (feature tables, data processing script and example chromatogram) have been deposited at <http://doi.org/10.5281/zenodo.4009045>

Author contributions

J.S conceptualized the study, developed the analytical methodology and sample preparation workflow. S.B and A.J carried out the experimental work and sample extractions. A.K carried out method validation and recovery studies. J.S developed the data processing workflow and wrote the code. S.B and J.S analyzed the data and wrote the manuscript. All authors performed scientific and grammatical revisions of the final draft.

Acknowledgments

The authors are grateful for the financial support received from the Danish Hydrocarbon Research & Technology Centre (DHRTC) under the Reservoir Fluids Characterization and Produced Water Management programs. The authors are also grateful for the donation of the samples from Total E&P Denmark. Furthermore, the authors kindly acknowledge the Danish Underground Consortium (Total, E&P Denmark, Noreco & Nordsøfonden) for granting permission to publish this work. Finally, the authors are grateful for fruitful and constructive feedback on the manuscript from Simon Ivar Andersen.

References

- Aitken CM, Jones DM, Larter SR. 2004. Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs. *Nature* 431:291–294. DOI: 10.1038/nature02922.
- Bakke T, Klungsøyr J, Sanni S. 2013. Environmental impacts of produced water and drilling waste discharges from the Norwegian offshore petroleum industry. *Marine Environmental Research* 92:154–169. DOI: 10.1016/j.marenvres.2013.09.012.
- Barros E V., Dias HP, Pinto FE, Gomes AO, Moura RR, Neto AC, Freitas JCC, Aquije GMF V., Vaz BG, Romão W. 2018. Characterization of Naphthenic Acids in Thermally Degraded Petroleum by ESI(–)-FT-ICR MS and ¹H NMR after Solid-Phase Extraction and Liquid/Liquid Extraction. *Energy & Fuels* 32:2878–2888. DOI: 10.1021/acs.energyfuels.7b03099.
- Barth T. 1991. Organic acids and inorganic ions in waters from petroleum reservoirs, Norwegian continental shelf: a multivariate statistical analysis and comparison with American reservoir formation waters. *Applied Geochemistry* 6:1–15. DOI: 10.1016/0883-2927(91)90059-X.
- Barth T, Borgund AE, Riis M. 1990. Organic acids in reservoir waters—Relationship with inorganic ion composition and interactions with oil and rock. *Organic Geochemistry* 16:489–496. DOI: 10.1016/0146-6380(90)90064-7.
- Barth T, Riis M. 1992. Interactions between organic acids anions in formation waters and reservoir mineral phases. *Organic Geochemistry* 19:455–482. DOI: 10.1016/0146-6380(92)90012-M.
- Bergfors SN, Schovsbo NH, Feilberg KL. 2020. Classification of and changes within formation water types in Danish North Sea Chalk: A study of the Halfdan, Dan, Kraka and Valdemar oil reservoirs. *Applied Geochemistry*:104702. DOI: 10.1016/j.apgeochem.2020.104702.
- Dudek M, Vik EA, Aanesen SV, Øye G. 2020. Colloid chemistry and experimental techniques for understanding fundamental behaviour of produced water in oil and gas production. *Advances in Colloid and Interface Science* 276:102105. DOI: 10.1016/j.cis.2020.102105.
- Durdevic P, Yang Z. 2018. Dynamic Efficiency Analysis of an Off-Shore Hydrocyclone System,

- Subjected to a Conventional PID- and Robust-Control-Solution. *Energies* 11:2379. DOI: 10.3390/en11092379.
- Durell G, Johnsen S, Røe-Utvik T, Frost T, Neff J. 2004. Produced Water Impact Monitoring In The Norwegian Sector Of The North Sea: Overview Of Water Column Surveys In The Three Major Regions. In: *SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production*. Society of Petroleum Engineers,. DOI: 10.2118/86800-MS.
- Endo S, Pfennigsdorff A, Goss K-U. 2012. Salting-Out Effect in Aqueous NaCl Solutions: Trends with Size and Polarity of Solute Molecules. *Environmental Science & Technology* 46:1496–1503. DOI: 10.1021/es203183z.
- Head IM, Gray ND, Larter SR. 2014. Life in the slow lane; biogeochemistry of biodegraded petroleum containing reservoirs and implications for energy recovery and carbon management. *Frontiers in Microbiology* 5. DOI: 10.3389/fmicb.2014.00566.
- Hsu CS, Hendrickson CL, Rodgers RP, McKenna AM, Marshall AG. 2011. Petroleomics: advanced molecular probe for petroleum heavy ends. *Journal of Mass Spectrometry* 46:337–343. DOI: 10.1002/jms.1893.
- Jiménez S, Andreozzi M, Micó MM, Álvarez MG, Contreras S. 2019. Produced water treatment by advanced oxidation processes. *Science of The Total Environment* 666:12–21. DOI: 10.1016/j.scitotenv.2019.02.128.
- Kiwi J, Lopez A, Nadtochenko V. 2000. Mechanism and Kinetics of the OH-Radical Intervention during Fenton Oxidation in the Presence of a Significant Amount of Radical Scavenger (Cl -). *Environmental Science & Technology* 34:2162–2168. DOI: 10.1021/es991406i.
- Kovalchik KA, MacLennan MS, Peru KM, Headley J V., Chen DDY. 2017. Standard method design considerations for semi-quantification of total naphthenic acids in oil sands process affected water by mass spectrometry: A review. *Frontiers of Chemical Science and Engineering* 11:497–507. DOI: 10.1007/s11705-017-1652-0.
- De Laat J, Le TG. 2006. Effects of chloride ions on the iron(III)-catalyzed decomposition of hydrogen peroxide and on the efficiency of the Fenton-like oxidation process. *Applied Catalysis B: Environmental* 66:137–146. DOI: 10.1016/j.apcatb.2006.03.008.
- Lin L, Jiang W, Chen L, Xu P, Wang H. 2020. Treatment of Produced Water with Photocatalysis: Recent Advances, Affecting Factors and Future Research Prospects. *Catalysts* 10:924. DOI: 10.3390/catal10080924.
- Liu B, Chen B, Zhang B, Song X, Zeng G, Lee K. 2021. Photocatalytic ozonation of offshore produced water by TiO₂ nanotube arrays coupled with UV-LED irradiation. *Journal of Hazardous Materials* 402:123456. DOI: 10.1016/j.jhazmat.2020.123456.

- Lofthus S, Almås IK, Evans P, Pelz O, Brakstad OG. 2018. Biodegradation in seawater of PAH and alkylphenols from produced water of a North Sea platform. *Chemosphere* 206:465–473. DOI: 10.1016/j.chemosphere.2018.05.006.
- Lyngbaek ME, Blidegn LH. 1991. Produced Water Management. In: *SPE Health, Safety and Environment in Oil and Gas Exploration and Production Conference*. Society of Petroleum Engineers,. DOI: 10.2118/23312-MS.
- Marshall AG, Rodgers RP. 2004. Petroleomics: The Next Grand Challenge for Chemical Analysis. *Accounts of Chemical Research* 37:53–59. DOI: 10.1021/ar020177t.
- Meldrum N. 1988. Hydrocyclones: A Solution to Produced-Water Treatment. *SPE Production Engineering* 3:669–676. DOI: 10.2118/16642-PA.
- Murphy RE, Schure MR, Foley JP. 1998. Effect of Sampling Rate on Resolution in Comprehensive Two-Dimensional Liquid Chromatography. *Analytical Chemistry* 70:1585–1594. DOI: 10.1021/ac971184b.
- Niu H, Lee K, Robinson B, Cobanli S, Li P. 2016. Monitoring and modeling the dispersion of produced water on the Scotian Shelf. *Environmental Systems Research* 5:19. DOI: 10.1186/s40068-016-0070-5.
- Norwegian Oil and Gas. 2012. *O85 – Norwegian Oil and Gas recommended guidelines for sampling and analysis of produced water*.
- OSPAR Commission. 2001. *Overview assessment of the implementation of OSPAR Recommendation 2001 / 1 for the management of produced water from offshore installations Offshore Industry Series*.
- Palacio Lozano DC, Thomas MJ, Jones HE, Barrow MP. 2020. Petroleomics: Tools, Challenges, and Developments. *Annual Review of Analytical Chemistry* 13:405–430. DOI: 10.1146/annurev-anchem-091619-091824.
- Pannekens M, Kroll L, Müller H, Mbow FT, Meckenstock RU. 2019. Oil reservoirs, an exceptional habitat for microorganisms. *New Biotechnology* 49:1–9. DOI: 10.1016/j.nbt.2018.11.006.
- Di Poi C, Costil K, Bouchart V, Halm-Lemeille M-P. 2018. Toxicity assessment of five emerging pollutants, alone and in binary or ternary mixtures, towards three aquatic organisms. *Environmental Science and Pollution Research* 25:6122–6134. DOI: 10.1007/s11356-017-9306-9.
- Røe TI, Johnsen S. 1996. Discharges of Produced Water to the North Sea. In: *Produced Water 2*. Boston, MA: Springer US, 13–25. DOI: 10.1007/978-1-4613-0379-4_3.
- Røe Utvik TI. 1999. Chemical characterisation of produced water from four offshore oil production platforms in the North Sea. *Chemosphere* 39:2593–2606. DOI: 10.1016/S0045-6535(99)00171-X.
- Samanipour S, Hooshyari M, Baz-Lomba JA, Reid MJ, Casale M, Thomas K V. 2019. The effect of extraction methodology on the recovery and distribution of naphthenic acids of oilfield produced

- water. *Science of The Total Environment* 652:1416–1423. DOI: 10.1016/j.scitotenv.2018.10.264.
- Samanipour S, Reid MJ, Rundberget JT, Frost TK, Thomas K V. 2020. Concentration and Distribution of Naphthenic Acids in the Produced Water from Offshore Norwegian North Sea Oilfields. *Environmental Science & Technology* 54:2707–2714. DOI: 10.1021/acs.est.9b05784.
- Saththasivam J, Loganathan K, Sarp S. 2016. An overview of oil–water separation using gas flotation systems. *Chemosphere* 144:671–680. DOI: 10.1016/j.chemosphere.2015.08.087.
- Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J. 2014. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science & Technology* 48:2097–2098. DOI: 10.1021/es5002105.
- Schymanski EL, Singer HP, Slobodnik J, Ipolyi IM, Oswald P, Krauss M, Schulze T, Haglund P, Letzel T, Grosse S, Thomaidis NS, Bletsou A, Zwiener C, Ibáñez M, Portolés T, de Boer R, Reid MJ, Onghena M, Kunkel U, Schulz W, Guillon A, Noyon N, Leroy G, Bados P, Bogialli S, Stipanichev D, Rostkowski P, Hollender J. 2015. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Analytical and Bioanalytical Chemistry* 407:6237–6255. DOI: 10.1007/s00216-015-8681-7.
- Sirtori C, Zapata A, Malato S, Agüera A. 2012. Formation of chlorinated by-products during photo-Fenton degradation of pyrimethanil under saline conditions. Influence on toxicity and biodegradability. *Journal of Hazardous Materials* 217–218:217–223. DOI: 10.1016/j.jhazmat.2012.03.017.
- Sørensen L, McCormack P, Altin D, Robson WJ, Booth AM, Faksness L-G, Rowland SJ, Størseth TR. 2019. Establishing a link between composition and toxicity of offshore produced waters using comprehensive analysis techniques – A way forward for discharge monitoring? *Science of The Total Environment* 694:133682. DOI: 10.1016/j.scitotenv.2019.133682.
- Strømgren T, Sørstrøm SE, Schou L, Kaarstad I, Aunaas T, Brakstad OG, Johansen Ø. 1995. Acute toxic effects of produced water in relation to chemical composition and dispersion. *Marine Environmental Research* 40:147–169. DOI: 10.1016/0141-1136(94)00143-D.
- Sundberg J, Feilberg KL. 2020. Characterization of heteroatom distributions in the polar fraction of North Sea oils using high-resolution mass spectrometry. *Journal of Petroleum Science and Engineering* 184:106563. DOI: 10.1016/j.petrol.2019.106563.
- Tang Y, Yin M, Yang W, Li H, Zhong Y, Mo L, Liang Y, Ma X, Sun X. 2019. Emerging pollutants in water environment: Occurrence, monitoring, fate, and risk assessment. *Water Environment Research* 91:984–991. DOI: 10.1002/wer.1163.
- The Danish Energy Agency. 2014. *Oil and Gas Production in Denmark 2014*.
- Thomas K V., Langford K, Petersen K, Smith AJ, Tollefsen KE. 2009. Effect-Directed Identification of

Naphthenic Acids As Important in Vitro Xeno-Estrogens and Anti-Androgens in North Sea Offshore Produced Water Discharges. *Environmental Science & Technology* 43:8066–8071. DOI: 10.1021/es9014212.