

RESEARCH

Open Access



# GC × GC–HRMS nontarget fingerprinting of organic micropollutants in urban freshwater sediments

Josephine S. Lübeck\* , Guilherme L. Alexandrino and Jan H. Christensen

## Abstract

**Background:** Sediments are sinks for organic micropollutants, which are traditionally analysed by gas chromatography–mass spectrometry (GC–MS). Although GC–MS and GC–tandem MS (MS/MS) are preferred for target screening, they provide only limited chromatographic resolution for nontarget screening. In this study, a comprehensive two-dimensional GC–high-resolution MS method (GC × GC–HRMS) was developed for nontarget screening and source identification of organic micropollutants in sediments from an urban channel and adjacent lake in Copenhagen, Denmark. The GC × GC–HRMS data were processed by pixel-based chemometric analysis using baseline subtraction, alignment, normalisation, and scaling before principal component analysis (PCA) of the pre-processed GC × GC–HRMS base peak ion chromatograms (BPCs). The analysis was performed to identify organic micropollutants of high abundance and relevance in the urban sediments and to identify pollution sources. Tentative identifications were based on match factors and retention indices and tagged according to the level of identification confidence.

**Results:** The channel contained both a significantly higher abundance of micropollutants and a higher diversity of compounds compared to the lake. The PCA models were able to isolate distinct sources of chemicals such as a natural input (viz., a high relative abundance of mono-, di- and sesquiterpenes) and a weathered oil fingerprint (viz., alkanes, naphthenes and alkylated polycyclic aromatic hydrocarbons). A dilution effect of the weathered oil fingerprint was observed in lake samples that were close to the channel. Several benzothiazole-like structures were identified in lake samples close to a high-traffic road which could indicate a significant input from asphalt or tire wear particles. In total, 104 compounds and compound groups were identified.

**Conclusions:** Several chemical fingerprints of different sources were described in urban freshwater sediments in Copenhagen using a pixel-based chemometric approach of GC × GC–HRMS BPCs. Various micropollutants of anthropogenic origin were identified. Tailored pre-processing and careful interpretation of the identification results is inevitable and still requires further research for an automated workflow.

**Keywords:** Chemical fingerprint, Sediment analysis, Source identification, GC × GC–HRMS, Pixel-based analysis, Prioritisation

## Background

Anthropogenic pollution of freshwater ecosystems via agricultural, industrial and urban activities is ubiquitous. The World Water Assessment Programme estimated in 2017 that approximately 80% of all wastewater, globally, is discharged into the environment without treatment [1]. However, environmental awareness is increasing and thus, the desire for a better understanding of what types

\*Correspondence: josephineluebeck@gmail.com  
Department of Plant and Environmental Sciences, Faculty of Science,  
University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg,  
Denmark

of pollutants are ending up in our environment. Anthropogenic micropollutants comprise numerous chemicals and their degradation products such as pharmaceuticals, detergents, pesticides and chemicals from consumer care products [2, 3]. These can be introduced to freshwater ecosystems via, e.g., household or industrial waste effluents; littering, road runoff and car exhaust. Compounds with low water solubility and high octanol/water partition coefficient ( $\log K_{OW}$ ) mostly deposit in sediments and are only released slowly to the water where they could cause adverse effects to the local fauna. The European Parliament and Council identified 45 priority substances in 2013 that every member state ought to monitor in surface waters at least once a year [4]. The list includes several heavy metals, pesticides, halogenated compounds, polycyclic aromatic hydrocarbons (PAHs) and phenols.

Gas chromatography (GC) coupled with mass spectrometry (MS) is the conventional solution for target analysis of volatile and semi-volatile persistent organic pollutants (POPs) in the environment [5]. The use of extracted ion chromatograms (EICs) facilitates the identification of known compounds; however, monitoring unknown chemicals or chemicals of emerging concern (CECs) that have previously not been included on any regulatory list, is a more challenging task. For a more comprehensive chemical impact assessment of our environment, suspect screening and nontarget screening (NTS) are used in combination with target analyses [2, 6]. The identification of unknown compounds in an NTS workflow is usually made by comparing experimental mass spectra with MS-libraries such as NIST for GC-MS with electron ionisation [7]. An adequate chromatographic resolution that provides mass spectra of the unidentified peaks free of chemical interferences is key to increase the reliability of the spectral matching. Therefore, GC-MS may not provide sufficient resolution for NTS of complex environmental matrices such as sediments.

Comprehensive two-dimensional GC with either low or high-resolution mass spectrometry detection ( $GC \times GC$ -(HR)MS) provides higher peak capacities than those obtained by one-dimensional GC both for target analysis of, e.g., PAHs, PAH derivatives and organochlorine pesticides [8]; and for NTS where more unambiguous identification of POPs has been obtained based on, e.g., high-resolution neutral loss of halogens, isotopic patterns and mass defect calculations [9, 10]. Additional applications of  $GC \times GC$ -(HR)MS for environmental purposes have also been reported [3, 11–14]. Data mining in NTS is still labour-intensive, and prioritisation and identification of thousands of peaks can be challenging, especially with large datasets [6, 15]. Methods for peak prioritisation in

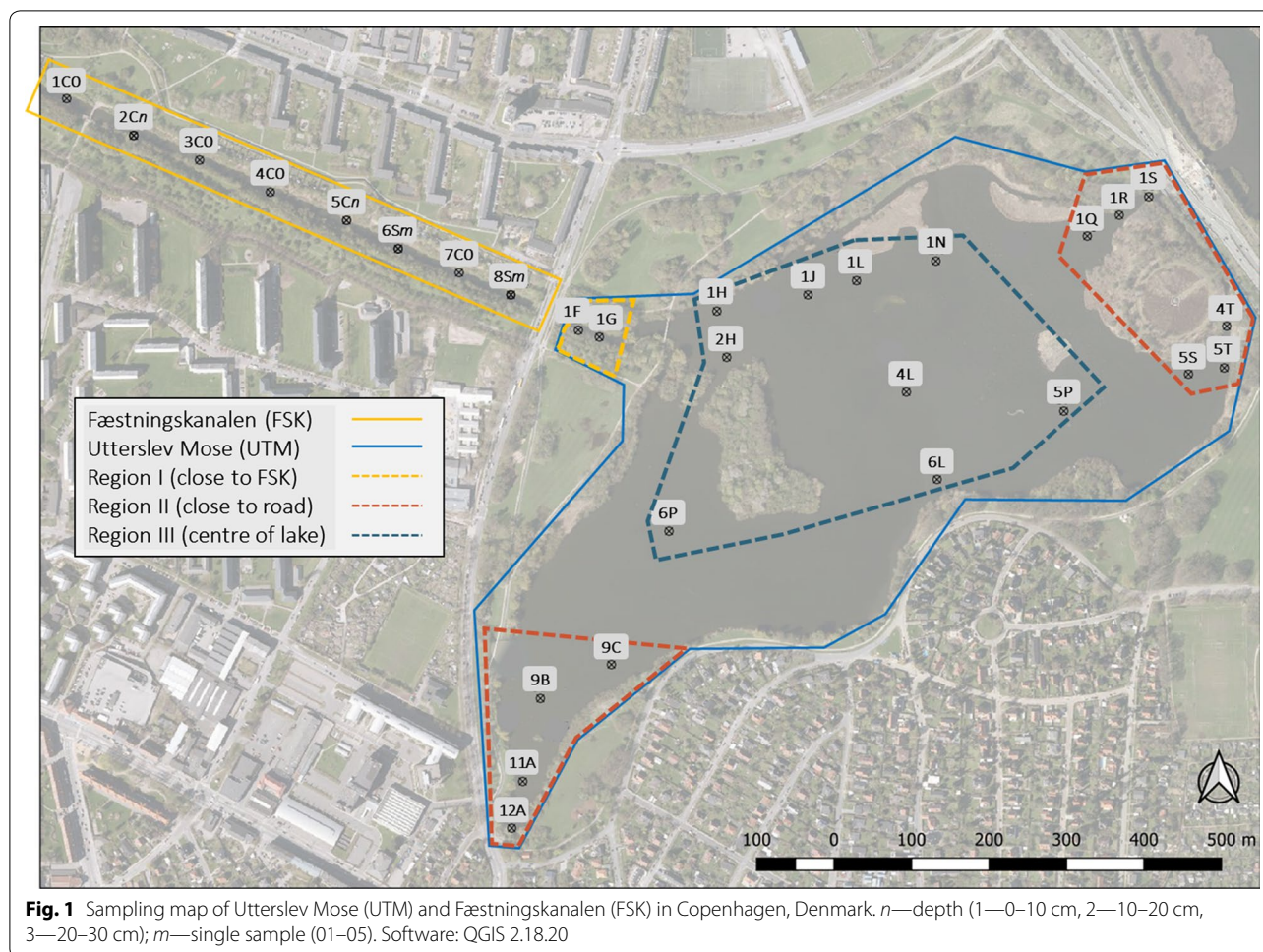
NTS have been described in the literature, e.g., by intensity, specific isotopic patterns or as part of a homologous series [6, 16], but few methods have been suggested for sample prioritisation. Even though chemometric tools, such as principal component analysis (PCA) [6] or hierarchical clustering analysis [17], have already been incorporated in NTS workflows to prioritise specific chemical profiles, applications in environmental monitoring are still limited, particularly for  $GC \times GC$ -(HR)MS. However, some benefits of chemometrics in NTS are apparent, e.g., the extraction of the most relevant chemical patterns or different sources of pollution can be visualised from the entire chemical fingerprint of the samples. In the so-called pixel-based approach, the data analysis is performed directly on the chromatographic pixels [18]. Therefore, the main variance in the data is displayed directly on the chromatographic space without prior peak-picking. The pixel-based approach facilitates the interpretation of structured chromatograms commonly found in  $GC \times GC$ -(HR)MS data, such as homologous series. While hierarchical clustering analysis is useful for assessing the overall chemical similarity among samples, the role of the individual variables, in this case, the chromatographic pixels that explain a similarity profile in the dataset cannot be evaluated. For example, a study by Alexandrino et al. successfully implemented the pixel-based analysis for forensic investigations of diesel spills in the environment based on  $GC \times GC$ -(HR)MS data [19].

This study aimed to characterise freshwater lake and channel sediments in an urban area (Copenhagen, Denmark) based on an NTS analytical workflow using  $GC \times GC$ -(HR)MS. First of all, the overall chemical variation between and within two sampling sites was investigated with pixel-based PCA. Second, a PCA for each sampling site intended to give a refined insight into the sources of pollution and chemical fingerprints. Subsequently, distinct samples were prioritised based on the pixel-based PCA for tentative compound identification.

## Materials and methods

### Sampling

Figure 1 shows the two sampling sites from the sampling campaign in September to November 2017: the lake Utterslev Mose (UTM) and the adjacent fortress channel (FSK as in Fæstningskanalen) in Copenhagen, Denmark. The lake is part of a protected nature park in the western part of Copenhagen and is, among others, fed by the fortress channel which surrounds the Danish capital. Sampling was performed with a Kajak sediment core sampler (KC Denmark A/S) with acrylic sample tubes (length: 47.5 cm, i.d.: 0.46 cm, wall thickness: 0.4 cm). Sediment samples (0–30 cm) were obtained from the lake (UTM) in a sampling grid of 50 × 50 m by collecting five



increments (i.e., five single core samples in the same grid) which were subsequently pooled to form one composite sample, e.g., 11A. Three types of sample were collected from eight sampling grids (80 × 10 m) at the fortress channel FSK: (i) four composite samples with each five increments such as for the lake samples (indicated by *CO* in Fig. 1); (ii) two composite samples (five increments each) that were pooled horizontally but not in vertical direction (indicated by *Cn* in Fig. 1, e.g., 2C1: 0–10; 2C2: 10–20; 2C3: 20–30 cm); (iii) in two sampling grids, five increment samples were kept separated, i.e., pooled vertically (0–30 cm), but not horizontally (indicated by *Sm* in Fig. 1, e.g., 6S01 to 6S05). A detailed sample overview is provided in Additional file 1: Table S1. After draining most of the surface water and proper mixing of the increments or single samples, representative mass reduction (between 5:1 and 2:1) was performed with a homemade mass reduction tool (stainless steel, Additional file 1: Figure S1) to approximately 200 g. Mass-reduced samples were transferred to Rilsan® bags and stored at 4 °C. In total, 19 and 28 samples were retrieved from FSK

and UTM, respectively. The lake was further divided into three regions: *Region I*—close to FSK, *Region II*—close to a road, and *Region III*—centre of the lake (Fig. 1).

### Chemicals

Methanol (MeOH) and ethanol (EtOH), both LC–MS grade (LC–MS Chromasolv®), were purchased from Sigma-Aldrich (St. Louis, MO, USA). Isooctane and dichloromethane (DCM) in HPLC grade were bought from Rathburn Chemicals (Walkerburn, Scotland), acetone from VWR-Chemicals (Radnor, PA, USA). MilliQ water (H<sub>2</sub>O) was obtained in-house with a type I ultrapure water purification system from ELGA-Veolia LabWater (High Wycombe, UK). Four spiking mixtures were prepared, containing aliphatic and aromatic acids, PAHs, steroids, pesticides, perfluorinated, deuterated and oxygen- and nitrogen-containing polycyclic aromatic compounds (*n* = 115 compounds, Additional file 1: Table S2).

### Sample preparation

Samples were extracted within one month after sampling. Pressurised liquid extraction (PLE) with an Accelerated Solvent Extractor (ASE-200 Dionex, USA) was applied. The 33-mL stainless-steel extraction cells were packed with two cellulose filters (GB-140, Advantec, MFS Inc., Japan) in the bottom. The sediment samples (7.5 g wet weight) were thoroughly mixed and ground with 22.5 g heat-treated (550 °C overnight) Ottawa sand (general purpose grade, Fisher Chemicals, Germany). The homogenised mixture was subsequently transferred to the extraction cell (average loss of  $4.17 \pm 0.76\%$  in weight inside the mortar). The remaining cell volume was filled with Ottawa sand and covered with a cellulose filter on top before the cell was closed. Two fractions were extracted in sequence from the same extraction cell: (i) polar fraction with MeOH:H<sub>2</sub>O (1:1) and (ii) nonpolar fraction with DCM:acetone (3:1). Extraction parameters for each fraction (i and ii) were as follows: pressure—1500 psi; temperature—(i) 50 °C/ (ii) 100 °C; pre-heating time—2 min, static time—10 min; flush volume—70%; purge time—60 s. Two static extraction cycles were performed into the same collection vial (60 mL amber-glass vials) per fraction. Six batches were run in total. Each batch contained one blank (extraction cell filled solely with Ottawa sand) and one randomly chosen sample replicate (denoted with an *R* in the label) per batch. Extracts were subsequently spiked with 200 µL of a deuterated standard mix (8 µg mL<sup>-1</sup> of each acenaphthylene-d<sub>8</sub>, dibenzothiophene-d<sub>8</sub>, naphthalene-d<sub>8</sub>, anthracene-d<sub>10</sub>, pyrene-d<sub>10</sub> and benz(a)anthracene-d<sub>12</sub> in isooctane). Extracts were stored at -20 °C. Residual water was removed from the nonpolar fraction by adding sodium sulphate (Na<sub>2</sub>SO<sub>4</sub>, Merck & Co., NJ, USA) in an excessive amount to the extracts, vortexing the mixture for 1 min and let the Na<sub>2</sub>SO<sub>4</sub> settle for 20 min. The supernatant was transferred to GC-vials and kept at -20 °C until analysis.

### Chemical analysis

Only the nonpolar fraction was analysed with GC × GC–HRMS. Extracts from FSK were diluted five times with the extraction solvent to adjust concentration differences between UTM and FSK samples as the total concentrations were higher in FSK. Facilitator samples from UTM (AnQC<sub>UTM</sub>) and FSK (AnQC<sub>FSK</sub>) were prepared by pooling equal volumes of eight and seven extracts from UTM or FSK, respectively (Additional file 1: Table S3). The facilitator samples were used during signal processing. Five batches containing ten randomly selected samples (including extraction replicates) and at least one of each AnQC<sub>UTM</sub> and AnQC<sub>FSK</sub> were analysed in each batch. A reference crude oil sample (1.25 mg mL<sup>-1</sup>), an n-alkane

series mix (Florida mix), the deuterated standard mixtures (50 ppb) and standard mixtures (Additional file 1: Table S2) at two nominal concentrations (2.5 and 10 ppb for the PAH mix in isooctane; 50 and 200 ppb for the other mixes) were analysed to facilitate the identification. An Agilent 7890B GC system (Agilent Technologies, Palo Alto, CA, USA) was modified with a secondary column oven and a Zoex<sup>®</sup> ZX2 cryogenic loop-type modulator (Zoex Corporation, Houston, TX, USA) and coupled to a 7200 Accurate Mass GC/QTOF mass spectrometer. The GC × GC separation was performed using a normal column set (nonpolar–polar), in which the first dimension (<sup>1</sup>D) consisted of a ZB-5 column (60 m, 0.25 mm i.d., 0.25 µm), while in the second dimension (<sup>2</sup>D) a ZB-50 column (1.5 m, 0.18 mm i.d., 0.18 µm) was used (Phenomenex, Torrance, CA, USA). A SilTite<sup>™</sup> µ-Union ferrule (SGE Analytical Science, Wetherill Park, Australia) connected the columns. Aliquots of 1.0 µL were injected in splitless mode with an inlet temperature of 300 °C. The primary oven temperature was ramped according to the following gradient: 60 °C held for 3.5 min, 7.5 °C min<sup>-1</sup> to 310 °C, and held for 15 min. The secondary oven temperature was operated at a constant temperature offset of +10 °C from the primary oven. Helium (≥99.9999%, AGA, Pullach, Germany) was used as carrier gas at a constant flow rate of 1.5 mL min<sup>-1</sup>. The modulation was performed using an independent cooling system that provided cold jets of N<sub>2</sub>(g) at approximately -70 °C, while the hot jets (pressure=20 psi) were produced with N<sub>2</sub>(g) heated at a constant temperature offset of +60 °C from the primary oven. The final 0.7 m of the <sup>1</sup>D was used as the dual-stage loop. The modulation period was set to 6.0 s, holding the hot jet for 0.5 s each modulation cycle. The electron ionisation (EI) source was operated at 230 °C with an emission current of 35 µA and 70 eV. The quadrupole time-of-flight (QTOF) mass analyser was employed in TOF mode (resolving power of 13,500 between 200 and 1000 amu) with a transfer line temperature of 310 °C and a quadrupole temperature of 150 °C. Mass calibration was performed in every batch after every five runs, using perfluorotributylamine (671.096 g mol<sup>-1</sup>) as the instrument mass calibration solution. Data acquisition was performed in the *m/z* range of 20–700 Da with 25 spectra s<sup>-1</sup>. Mass-Hunter (B.07.00, Agilent Technologies) was used to control the instrument.

### Data treatment

#### Pre-processing

The data files were converted to netCDF files using the AIA File Translator programme from Agilent which bins the raw data in ±0.025 Da bins. An in-house script was written to import netCDF files into

Matlab R2017b (MathWorks, Natick, MA, USA). The script applied a filtering step that excluded the most dominant (high-resolution)  $m/z$  values derived from column bleeding (Additional file 1: Table S4). The two-dimensional base peak ion chromatograms (2D-BPCs) ( ${}^1t_R \times {}^2t_R$ ) were extracted, where  ${}^1t_R$  and  ${}^2t_R$  are the retention times in the first ( ${}^1D$ ) and second ( ${}^2D$ ) dimension, respectively. Each 2D-BPC was phase-corrected to account for rigid retention time shifts due to the modulation.

Further, due to differences in total concentration and sample complexity, the dataset was divided according to the two sampling sites. For alignment, a 2D correlation optimised warping algorithm (2D-COW) was used [20], where each 2D-BPC was aligned to the target BPC from the corresponding AnQC (FSK or UTM) obtained from the third batch (middle of the sequence). The optimal warping parameters, segment length and slack (viz., how much each segment is allowed to change in the alignment) were obtained with an in-house script in Matlab that performs a grid search in the parameters space, herein considering both chromatographic dimensions. The two aligned datasets were individually unfolded into row-vectors that resulted in a data matrix  $\mathbf{D}_i(K, L)$ ,  $i = \text{UTM}$  or  $\text{FSK}$ , where  $K$  is the number of objects (i.e., samples, replicates and AnQC samples) in sampling site  $i$  and  $L$  the number of variables or pixels that represent the aligned BPCs. Blank-subtraction was performed by removing the peaks that were present in the aligned BPC of the blanks suggesting column bleeding, instrumental contamination and contamination from the sample preparation. Finally,  $\mathbf{D}_i$  was normalised to the unitary Euclidian norm to focus the subsequent pixel-based analysis on the relative composition and reducing the concentration effects markedly. The normalisation also aims to remove the minor variation of injection volume between samples and to reduce the effects of variations in the instrument sensitivity along with the analysis of the batches [21]. Small concentration effects will be retained even after this normalisation as the amount of noise in each BPC would affect the data analysis. This normalisation procedure corresponds to dividing each concentration (in a targeted quantitative study) by the sum of the concentrations. The effect is that any prior normalisation to, e.g., dry weight, organic matter content or single internal standards would be removed again and therefore it makes little sense to pre-normalise to, e.g., the dry weight. The BPCs of the reference crude oil, Florida mix, blanks and standards were not included in  $\mathbf{D}_i$ ; they were used for quality assurances and identification and not for the modelling part (see below).

#### **Pixel-based analysis and weighted-principal component analysis (WPCA)**

The analyses were performed with WPCA [19]. The variation that is unrelated to the chemical composition (e.g., instrumental noise, residual retention time shifts, column bleeding and peak saturation) negatively affects the quality of PCA models. Therefore, each BPC in  $\mathbf{D}_i$  was weighted by dividing each row of  $\mathbf{D}_i$  element-wise by a vector  $\mathbf{w}$  (defined as the pixel-by-pixel relative analytical standard deviations (RSD) of the BPC exclusively from the AnQCs). The weighting down-scales pixels noise regions (e.g., electronic and chemical noise) and regions of the BPC where the alignment is poor (e.g., fronting and tailing sections), and up-scales regions that contain chemical information (peak regions) [22]. The *global model* aimed to extract the main chemical variation that distinguishes UTM and FSK and to compare the chemical heterogeneity within each sampling site. First,  $\mathbf{D}_{\text{FSK}}(M, L)$  and  $\mathbf{D}_{\text{UTM}}(N, L)$  were column-wise combined into an augmented matrix  $\mathbf{D}(O, L)$ , where  $O = M + N$ . The alignment of the peaks in  $\mathbf{D}$  after the combination of the individual  $\mathbf{D}_i$  matrices was evaluated beforehand. The dataset in the *global model* comprised an unfolded matrix of  $62 \times 35,000$  (samples  $\times$  data points) combining the two datasets of UTM and FSK, and included the facilitator samples  $\text{AnQC}_{\text{UTM}}$  and  $\text{AnQC}_{\text{FSK}}$  ( $6 \times 35,000$  and  $5 \times 35,000$ , respectively). Subsequently, the more detailed chemical variability occurring within each of the two sampling sites (UTM and FSK) was assessed through *local models* fitted for  $\mathbf{D}_{\text{FSK}}$  ( $23 \times 35,000$ ) and  $\mathbf{D}_{\text{UTM}}$  ( $39 \times 35,000$ ), respectively. All models were fitted on the mean-centred data. Only principal components (PCs) that contain chemical information rather than noise were further evaluated. The number of significant components in the PCA models was determined based on the extraction of only the PCs explaining predominantly systematic variation and not residual retention shift. Residual retention shifts would be clearly visible in the loadings as the loading coefficients for the front and tailing parts of single peaks would have a different sign (negative and positive loadings coefficients). In such cases, the chemical information of the PC would be confounded by shifts and changes in peak shape.

#### **Compound identification**

The interpretation of the models in an environmental context requires the (tentative) identification of peaks expressed in the PC loadings. Additionally, the score plot for each PC can be utilised to select samples in which the corresponding chromatographic pattern obtained from the PC loadings is more evident, i.e., the samples that are projected in the corners of the score plot for a particular PC. Selected raw data files were converted to *\*.mzXML*

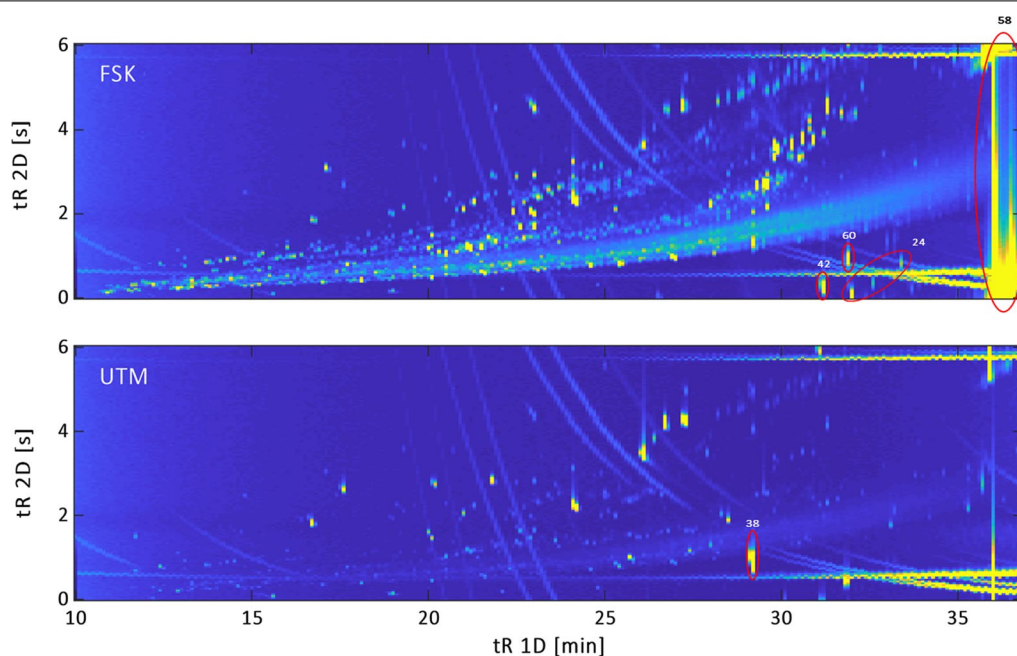
files using ProteoWizard (v 3.0.19140) and subsequently imported to Matlab. The data was re-folded into the original 2D-structure. Next, mass spectra at the maximum height of each identified peak in the corresponding 2D-BPC were extracted and organised in individual text files which were submitted to NIST14<sup>®</sup> MS library (Gaithersburg, MD, USA). The cut-off for a mass spectra matching was a match factor (MF) of  $\geq 800$ —all hits that were below that value were not considered for identification. The hit with the highest MF was selected for identification, and the results were organised in a list with all tentatively identified compounds. All the steps in Matlab were performed using in-house scripts.

MassHunter's software Unknowns Analysis (B.07.00) was used as an additional tool for compound identification to verify the results from the Matlab workflow. Parameters are described in Additional file 1: Table S5. Additionally, compounds that were part of the spiking mixtures were targeted. Identified compounds that were found with Unknowns Analysis and the in-house Matlab workflow were reported here, including structural identifiers, experimental and literature retention indices (RI, based on *n*-alkanes C10-C26 in the reference oil and Florida mix, Additional file 1: Table S6), and identification confidence levels (with Level 1—confirmed structure with reference standard; Level 4—unequivocal molecular formula) [23]. If the experimental and literature RI were different by more than 50 units, the confidence level was set down to Level 4.

## Results and discussion

### Screening of sample extracts

Substantial concentration differences were observed between UTM and FSK samples. In Fig. 2, the 2D-BPCs of facilitator samples for FSK (AnQC<sub>FSK</sub>) and UTM (AnQC<sub>UTM</sub>) are compared. The 2D-BPC of AnQC<sub>FSK</sub> contains >1000 peaks while 2D-BPC of AnQC<sub>UTM</sub> is much less populated, which demonstrates that the overall concentration of compounds in the two combined extracts is significantly higher in the fortress channel (FSK) compared to the lake (UTM). Some peaks in AnQC<sub>FSK</sub> expressed wraparound after 30 min, viz., peaks that should have retention in <sup>2</sup>D higher than the modulation period of 6 s. Fortunately, these wraparounds did not co-elute in <sup>2</sup>D with other peaks. Moreover, a detector overload can be seen in the 2D-BPC of AnQC<sub>FSK</sub> (#58, large yellow blob after 35 min in Fig. 2). This peak or cluster of peaks (which is difficult to assess due to the detector overload) was identified as phthalic acid esters (Table 1) because of the major fragment and characteristic base peak at *m/z* 149.0235. The most common member of the class of phthalates is bis(2-ethylhexyl) phthalate (DEHP); it is often found in environmental samples due to its excessive overuse in countless products and has been listed as a priority substance by the EU Parliament [4, 24]. Several potential candidates of phthalates are listed in Additional file 2. The detector saturation was observed only in a few FSK samples but in none from UTM or in the extraction blanks, indicating that



**Fig. 2** 2D-BPC of the two facilitator samples from FSK and UTM. Highlighted peaks: 24—DDD, 38—diphenyl sulfone, 42—fluoranthene, 58—phthalate acid esters, 60—pyrene (Table 1)

**Table 1** Tentatively identified compounds from both sampling sites (UTM and FSK)

#	Compound/compound group	Molecular formula	Monoisotopic mass [Da]	Mass error [ppm] <sup>a</sup>	<sup>1</sup> D t <sub>R</sub> [min]	ID level <sup>b</sup>
1	2-Mercaptobenzothiazole	C <sub>7</sub> H <sub>5</sub> NS <sub>2</sub>	166.9863	11.0	29.5	1
2	Acenaphthylene	C <sub>12</sub> H <sub>8</sub>	152.0626	10.5	21.9	1
3	Acenaphthene	C <sub>12</sub> H <sub>10</sub>	154.0783	13.3	22.5	1
4	Acetylides	C <sub>18</sub> H <sub>34</sub>	250.2661	7.0	25.85	4
5	Alkanes	C <sub>n</sub> H <sub>2n+2</sub>		7.2		3
6	Alkylphenylketone (phenacylidene diacetate)	C <sub>12</sub> H <sub>12</sub> O <sub>5</sub>	236.0685	13.1	11.7	4
7	Alkylphenols			5.7	24.1–25.3	3
8	Anthracene	C <sub>14</sub> H <sub>10</sub>	178.0783	5.3	27.3	1
8.1	Anthracenes, C1-	C <sub>15</sub> H <sub>12</sub>	192.0939	7.6	28.9–29.1	2
8.d	Anthracene derivative	C <sub>15</sub> H <sub>14</sub>	194.1096	8.0	28.0	3
9	Azulene	C <sub>10</sub> H <sub>8</sub>	128.0626	5.5	18.1	4
9.3	Azulene, C3-	C <sub>13</sub> H <sub>14</sub>	170.1096	5.6	23.1	4
10.3	Benzenes, C3-	C <sub>9</sub> H <sub>12</sub>	120.0939	11.5	11.7–13.1	3
10.4	Benzenes, C4-	C <sub>10</sub> H <sub>14</sub>	134.1096	6.2	13.1–16.0	3
10.5	Benzenes, C5-	C <sub>11</sub> H <sub>16</sub>	148.1252	8.3	14.8–17.7	3
10.6	Benzenes, C6-	C <sub>12</sub> H <sub>18</sub>	162.1409	7.6	16.7–20.1	3
10.7	Benzenes, C7-	C <sub>13</sub> H <sub>20</sub>	176.1565	10.2	20.0	3
10.8	Benzenes, C8-	C <sub>14</sub> H <sub>22</sub>	190.1722	5.9	21.8–21.9	3
11	Benzoic acid ester (methyl benzoate)	C <sub>8</sub> H <sub>8</sub> O <sub>2</sub>	136.0511	9.8	14.7	3
12	Benzothiazole	C <sub>7</sub> H <sub>5</sub> NS	135.0143	10.1	17.6	3
13	Benzoyl hydrazine	C <sub>7</sub> H <sub>8</sub> N <sub>2</sub> O	136.0524	14.2	14.8	4
14	Benzyl 2-chloroethyl sulfone	C <sub>9</sub> H <sub>11</sub> ClO <sub>2</sub> S	218.0168	14.5	20.2	4
15	Benzyl isocyanate	C <sub>8</sub> H <sub>7</sub> NO	133.0528	17.0	15.4	3
16	Benzyl thiocyanate	C <sub>8</sub> H <sub>7</sub> NS	149.0299	6.2	20.5	4
17	Benzylidenebenzylamine	C <sub>14</sub> H <sub>13</sub> N	195.1048	10.8	26.4	3
18	Biphenyl	C <sub>12</sub> H <sub>10</sub>	154.0783	10.1	20.5	3
18.1	Biphenyls, C1-	C <sub>13</sub> H <sub>12</sub>	168.0939	10.9	22.3–23.7	4
18.2	Biphenyls, C2-	C <sub>14</sub> H <sub>14</sub>	182.1096	10.7	24.0–24.2	3
19	Butylated hydroxytoluene	C <sub>15</sub> H <sub>24</sub> O	220.1827	6.0	22.6	2
20	Calamenes	C <sub>15</sub> H <sub>22</sub>	202.1722	4.9	21.8	4
21	Chloronaphthalenes	C <sub>10</sub> H <sub>7</sub> Cl	162.0236	9.4	20.6	3
22	Cyclic organosulphur			9.8		2
23	Cycloalkane	C <sub>13</sub> H <sub>26</sub>	182.2035	8.0	17.4	3
24	DDD	C <sub>14</sub> H <sub>10</sub> C <sub>14</sub>	317.9537	8.6	32.0–32.6	3
25	DDMS	C <sub>14</sub> H <sub>11</sub> C <sub>13</sub>	283.9926	5.8	31.2	4
26	Decalin	C <sub>10</sub> H <sub>18</sub>	138.1409	7	14.0	3
26.1	Decalins, C1-	C <sub>11</sub> H <sub>20</sub>	152.1565	8.5	15.2	3
26.2	Decalins, C2-	C <sub>12</sub> H <sub>22</sub>	166.1722	6.7	16.3–17.6	3
27	Dialkyl ethers	C <sub>n</sub> H <sub>2n+4</sub> O		6.1		4
28	Dibenzofuran	C <sub>12</sub> H <sub>8</sub> O	168.0575	1.9	23.0	1
29	Dibenzopyrans	C <sub>13</sub> H <sub>10</sub> O	182.0732	9.4	24.7–24.9	3
30	Dibenzothiophene	C <sub>12</sub> H <sub>8</sub> S	184.0347	3.6	26.8	1
30.1	Dibenzothiophenes, C1-	C <sub>13</sub> H <sub>10</sub> S	198.0503	8.5	28.7–28.9	2
30.2	Dibenzothiophenes, C2-	C <sub>14</sub> H <sub>12</sub> S	212.0660	8.1	29.6–30.4	2
31	Dibenzylamine	C <sub>14</sub> H <sub>15</sub> N	197.1204	2.8	26.1	3
32	Dicarboxylic acid derivatives	C <sub>16</sub> H <sub>30</sub> O <sub>4</sub>	286.2144	15.4	15.8	4
33	Dichlorobenzenes	C <sub>6</sub> H <sub>4</sub> Cl <sub>2</sub>	145.9690	6.2	12.7–13.4	3
34	Dichlorobisphenyl	C <sub>13</sub> H <sub>10</sub> Cl <sub>2</sub>	236.0160	7.8	28.1	3
35	Dichloroisocyanatobenzenes	C <sub>7</sub> H <sub>3</sub> Cl <sub>2</sub> NO	186.9592	7.0	19.1–19.4	4

**Table 1 (continued)**

#	Compound/compound group	Molecular formula	Monoisotopic mass [Da]	Mass error [ppm] <sup>a</sup>	<sup>1</sup> D t <sub>R</sub> [min]	ID level <sup>b</sup>
36	Dichloronaphthalenes	C <sub>10</sub> H <sub>6</sub> Cl <sub>2</sub>	195.9847	9.7	23.7–24.1	3
37.3	Dihydronaphthalenes, C3-	C <sub>13</sub> H <sub>16</sub>	172.1252	5.2	20.0–20.3	3
38	Diphenyl sulfone	C <sub>12</sub> H <sub>10</sub> O <sub>2</sub> S	218.0402	12.2	29.2	4
39	Diterpenes			5.5	28.9–31.3	3
39.1	10,18-Bisnorabieta-8,11,13-triene	C <sub>18</sub> H <sub>26</sub>	242.2035	6.0	30.6	3
39.2	10,18-Bisnorabieta-8,11,13-triene, C1-	C <sub>19</sub> H <sub>28</sub>	256.2191	7.4	32.1	4
40	Fatty acid methyl esters			8.4	26.6–31.2	3
41	Fatty alcohol (cis-9-eicosen-1-ol)	C <sub>20</sub> H <sub>40</sub> O	296.3079	9.8	28.7	4
42	Fluoranthene	C <sub>16</sub> H <sub>10</sub>	202.0783	2.2	31.9	1
43	Fluorene	C <sub>13</sub> H <sub>10</sub>	166.0783	12.3	24.1	1
43.1	Fluorenes, C1-	C <sub>14</sub> H <sub>12</sub>	180.0939	7.8	25.9–26.4	2
43.2	Fluorenes, C2-	C <sub>15</sub> H <sub>14</sub>	194.1096	6.8	27.7–28.2	3
44.1	Indanes, C1-	C <sub>10</sub> H <sub>12</sub>	132.0939	10.2	16.0	3
44.2	Indanes, C2-	C <sub>11</sub> H <sub>14</sub>	146.1096	10.3	16.7–18.0	3
44.3	Indanes, C3-	C <sub>12</sub> H <sub>16</sub>	160.1252	10.5	18.3–19.4	3
44.5	Indanes, C5-	C <sub>14</sub> H <sub>20</sub>	188.1565	10.1	22.9	3
45.6	Indanones, C6-	C <sub>18</sub> H <sub>24</sub> O	256.1827	6.2	31.4	3
46.2	Indenes, C2-(1-ethylidene-1H-indene)	C <sub>11</sub> H <sub>10</sub>	142.0783	12.3	19.3	3
46.4	Indenes, C4-	C <sub>13</sub> H <sub>16</sub>	172.1252	10.5	21.5	4
47	Monoterpenes	C <sub>15</sub> H <sub>24</sub>	204.1878	6.2	20.8–22.9	3
48	Naphthalene	C <sub>10</sub> H <sub>8</sub>	128.0626	5.5	16.8	1
48.1	Naphthalenes, C1-	C <sub>11</sub> H <sub>10</sub>	142.0783	7.7	18.9–19.0	2
48.2	Naphthalenes, C2-	C <sub>12</sub> H <sub>12</sub>	156.0939	8.4	20.8–21.9	2
48.3	Naphthalenes, C3-	C <sub>13</sub> H <sub>14</sub>	170.1096	6.7	21.9–23.7	2
48.4	Naphthalenes, C4-	C <sub>14</sub> H <sub>16</sub>	184.1252	4.2	24.2–26.4	2
48.d	Naphthalene derivatives			9.9		4
49.1	Naphthothiophenes, C1-	C <sub>13</sub> H <sub>10</sub> S	198.0503	6.2	28.2	3
49.2	Naphthothiophenes, C2-	C <sub>14</sub> H <sub>12</sub> S	212.0660	6.5	30.3	4
50	Nonyl pentyl sulfite	C <sub>14</sub> H <sub>30</sub> O <sub>3</sub> S	278.1916	5.8	13.9	4
51	PCBs	C <sub>12</sub> H <sub>4</sub> Cl <sub>6</sub>	357.8444	6.5	33.2–34.5	3
52	Pentafluorinated depsides	C <sub>14</sub> H <sub>7</sub> F <sub>5</sub> O <sub>2</sub>	302.0366	6.2	27.3	4
53	Phenylene	C <sub>13</sub> H <sub>10</sub>	166.0783	6.3	24.2	3
54	Phenanthrene	C <sub>14</sub> H <sub>10</sub>	178.0783	1.4	27.2	1
54.1	Phenanthrenes, C1-	C <sub>15</sub> H <sub>12</sub>	192.0939	6.0	28.8–29.2	2
54.2	Phenanthrenes, C2-	C <sub>16</sub> H <sub>14</sub>	206.1096	5.1	30.2–30.8	2
54.4	Phenanthrenes, C4-	C <sub>18</sub> H <sub>18</sub>	234.1409	3.6	32.8	3
54.d	Phenanthrene derivatives			11.3		4
55	Phenylacetaldehydes	C <sub>9</sub> H <sub>10</sub> O	134.0732	12.4	13.5	3
56	Phenylglyoxal	C <sub>8</sub> H <sub>6</sub> O <sub>2</sub>	134.0368	13.3	11.8	4
57	Phenyl-naphthalenes	C <sub>16</sub> H <sub>12</sub>	204.0939	7.1	29.7–29.8	4
58	Phthalic acid ester			5.4	35.5–36.7	3
59	Polycyclic hydrocarbons			6.8		4
60	Pyrene	C <sub>16</sub> H <sub>10</sub>	202.0783	6.7	31.9	1
61	Sesquiterpenes			6.6	20.0–26.8	3
62	Styrenes	C <sub>10</sub> H <sub>12</sub>	132.0939	10.6	14.5–15.3	3
63	Tetrachloroethane	C <sub>2</sub> H <sub>2</sub> Cl <sub>4</sub>	165.8911	6.4	10.5	3
64	Tetralin	C <sub>10</sub> H <sub>12</sub>	132.0939	7.6	16.2	3
64.1	Tetralins, C1-	C <sub>11</sub> H <sub>14</sub>	146.1096	11.5	18.1–18.9	3
64.2	Tetralins, C2-	C <sub>12</sub> H <sub>16</sub>	160.1252	10.8	19.3–20.8	3



**Table 1 (continued)**

#	Compound/compound group	Molecular formula	Monoisotopic mass [Da]	Mass error [ppm] <sup>a</sup>	<sup>1</sup> D t <sub>R</sub> [min]	ID level <sup>b</sup>
64.3	Tetralins, C3-	C <sub>13</sub> H <sub>18</sub>	174.1409	7.2	20.1–22.4	3
64.4	Tetralins, C4-	C <sub>14</sub> H <sub>20</sub>	188.1565	11.7	19.9–22.4	3
64.d	Tetralins derivative (versalide)	C <sub>18</sub> H <sub>26</sub> O	258.1984	3.7	29.6	3
65	Thiophene, 3-methyl-2-3,7,11-trimethyldodecyl	C <sub>20</sub> H <sub>36</sub> S	308.25377	9.0	31.4	3
66	Unsaturated hydrocarbons			8.3		3
67	Vanillins	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	152.0473	13.4	20.7–21.1	3

Excel sheet in Additional file 2 with more detailed information, incl. RI and chemical identifiers. Numbering (#) related to the loading plots in Figs. 4 and 7

<sup>a</sup> If several compounds were identified for a compound group, the average of the mass error was calculated

<sup>b</sup> Confidence ID levels from Schymanski et al. [23]. Retention time (<sup>1</sup>D t<sub>R</sub>) and index (RI) were extracted from the facilitator samples AnQC<sub>FSK</sub> and AnQC<sub>UTM</sub> and calculated based on *n*-alkane elution in the reference oil and Florida mix (Additional file 1: Table S6). If the calculated RI deviated ≥ 100 from the NIST library, the confidence ID level was set down to 4

the signal originated from the sample and not the sample preparation procedure (*cf.* Additional file 1: Figure S2). As the detector saturation would have exacerbated the pixel-based analysis, the cut-off for modelling was set to ≤ 35 min.

A qualitative target screening was performed for the spiked compounds listed in Additional file 1: Table S2 in the GC × GC–HRMS chromatograms of standard mixtures and the two facilitator samples from the third batch. In the standard mixture, 67 out of 109 compounds were detected; 25 out of the 67 were also found in the facilitator samples (Additional file 1: Table S2). Among the 42 compounds that were not detected with GC × GC–HRMS were aromatic and fatty acids, steroids, compounds of either high polarity or low volatility and (alkylated) four-, five- and six-ring PAHs. Possible explanations for not detecting these compounds can be, e.g., degradation at high temperatures, the inability to transfer the molecules into the gas-phase inside the liner, or the inability to elute strongly retained compounds due to the lower oven temperature limit determined by the secondary polar column compared to GC–MS methods with nonpolar columns. Additional analytical platforms need to be applied to determine these compounds such as liquid or supercritical fluid chromatography with electrospray ionisation MS; however, this was outside the scope of this study. Some of the compounds that were detected with GC × GC–HRMS represent groups of isomers with the same monoisotopic mass and molecular formula, and thereby enabled the identification of groups of these isomers in other samples.

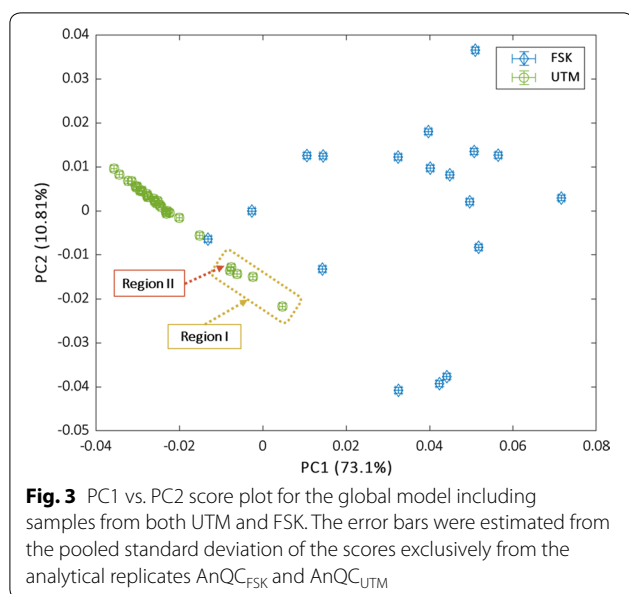
#### Pixel-based analysis

The enhanced peak capacity and sensitivity of GC × GC–HRMS allows the separation of thousands of compounds with different physicochemical properties in such complex environmental samples. Compound identification,

however, can be cumbersome still, because of the often overwhelming number of peaks and large datasets in GC × GC–HRMS. Prioritisation is often unavoidable and is based, for example, on signal intensity, peaks with a specific isotopic pattern or mass defect, to name a few [6, 16]. Often, some samples do not add meaningful information, especially when many samples are collected for spatial and temporal investigations. The pixel-based PCA provides information on the highest variation in a dataset, and thus, helps to focus the identification only on the samples with the highest variation and unique chemical fingerprints. Furthermore, the positive and negative signals in a loading plot can be used to identify peaks of high relevance (Table 1, Excel sheet in Additional file 2).

#### Global model

The global model includes all measured samples from both sampling sites and describes the overall variation within and between the two sampling sites. The score plot in Fig. 3 is a projection of the samples in the WPCA model (Additional file 1: Figure S3 for loading plots). Thus, the chemical similarities between pair or group of samples can be assessed while comparing the distances of their coordinates in this variable-reduced space spanned by the PCs. The first PC explains 73.10% of the total variance, whereas PC2 describes 10.81% (Fig. 3). Samples from FSK showed a more considerable variation along the PCs subspace compared to the UTM samples, which also demonstrates that the FSK samples contain a more substantial chemical heterogeneity across the sampling site. In general, the samples were separated in the WPCA model according to sampling location. However, there is an overlap along PC1 between UTM samples collected close to the outlet of the channel (*Region I*) and close to the road (*Region II*) (Fig. 3). A reasonable hypothesis is that the samples collected in these specific locations of the lake are affected by chemical inputs from FSK and



the urban areas delimited by *Region II*. In contrast, the UTM samples from the centre of the lake (*Region III*) are less affected by the chemical inputs from both FSK and *Region II*, due to dilution in the lake.

In summary, the global WPCA model was able to show that (i) FSK and UTM sampling sites contain distinct chemical fingerprints and (ii) chemical inputs to UTM may come from FSK (*Region I*) and the urban areas delimited by *Region II*. To assess the chemical composition of the samples from each sampling site individually, and to elucidate the contamination sources, local WPCA models for FSK and UTM sample sets were calculated. The local WPCA models were used to prioritise a subset of samples within each site in order to reduce the identification workload.

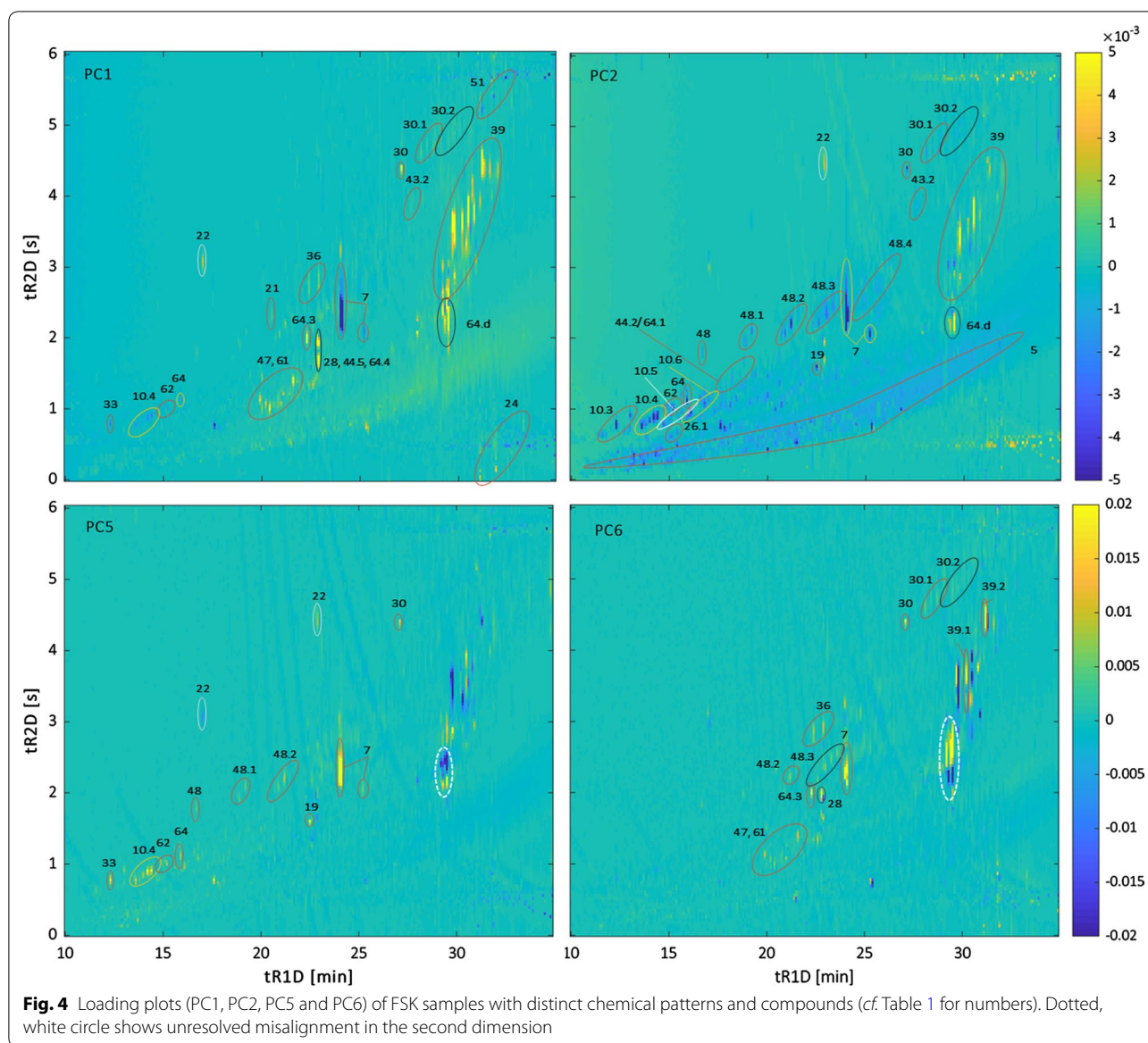
#### Local models—channel (FSK) site

The local model of the FSK sampling site explained 91.27% of the total variation and was built using six PCs. Chemical interpretation of the loading coefficients of PC1 to PC6 was performed to assess chemically relevant patterns and the presence of modelling artefacts (Fig. 4). For example, a negative score in the PC1 loading, such as the dark blue peak #7 in Fig. 4, and a negative score in the scores plot (Fig. 5) indicates that this particular sample (2C1) has a high relative abundance (due to the nature of the normalisation) of that compound compared to the samples with high positive PC1 scores. PC3 and PC4 loadings majorly describe residual retention time shifts that could not be fully removed during pre-processing (Additional file 1: Figure S4). Therefore, they do not

explain relevant chemical information from this site and were not discussed further.

Chemical interpretation of the PC1 loading (explained variance of 44.80%) together with the BPCs of the particular samples with the highest negative and highest positive scores (2C1 and 2C3, respectively, Fig. 5) revealed a higher relative input of natural chemically occurring compounds as a function of the depth in the sediment. An increase in complexity and the number of peaks is noticeable with increasing depth when investigating the GC × GC–HRMS chromatograms of samples 2C1, 2C2 and 2C3 (Fig. 6). Many peaks were identified as monoterpenes (#47), sesquiterpenes (#61) or diterpenes (#39). These compounds are commonly derived biosynthetically from diverse plants, fungi and animals [25]. Because of the nature of the extraction solvent mixture (DCM:acetone [3:1]) and the absence of an in-cell adsorbent, many mid-polar components were also extracted, such as diterpenes. Typically, these naturally occurring components are removed in sample extraction and cleanup [14, 26]. Here, terpenes were associated with a natural input chemical fingerprint.

The PC2 loading (Fig. 4) describes a weathered oil fingerprint (blue peaks with a negative score) [27, 28] because of the occurrence of many hydrocarbons, naphthenes and alkylated mono- and polycyclic aromatic hydrocarbons, some of which were confirmed with standards (Level 1 in Table 1, Additional file 1: Table S2). They indicate a continuous input of mineral oil products in this particular region of the fortress channel (negatively scoring samples 1C to 2C). The top-layer (2C1, 0–10 cm) scores positively in PC2, thus, there is relatively less of these oil components within the entire chromatographic fingerprint of these samples. The retained oil compounds were weathered and potentially biodegraded in the lower sediment levels (2C2 to 2C3, 10–30 cm, Fig. 6) seen by the many alkylated compounds such as the C4- to C6-benzenes (#10.4–10.6), naphthalenes (#48.1–48.4), PAHs and dibenzothiophenes (#30.1–30.2). Alternatively, the higher degradation of the natural products relatively to the oil compounds in the deeper layers (2C2 and 2C3) can also be a reasonable hypothesis to explain the higher input of oil compounds within the chromatograms of these samples. Alkylphenols (#7) and butylated hydroxytoluene (#19) also occurred with negative PC2 coefficients (Fig. 5). They have a wide variety of applications in consumer products such as detergents or cleaning products; butylated hydroxytoluene is known as an antioxidant and is widely used in fuels to prevent oxidation. The most common alkylated phenol is nonylphenol which is primarily used for the production of nonylphenol ethoxylate surfactants or found as a degradation product of them. Alkylated phenols are defined

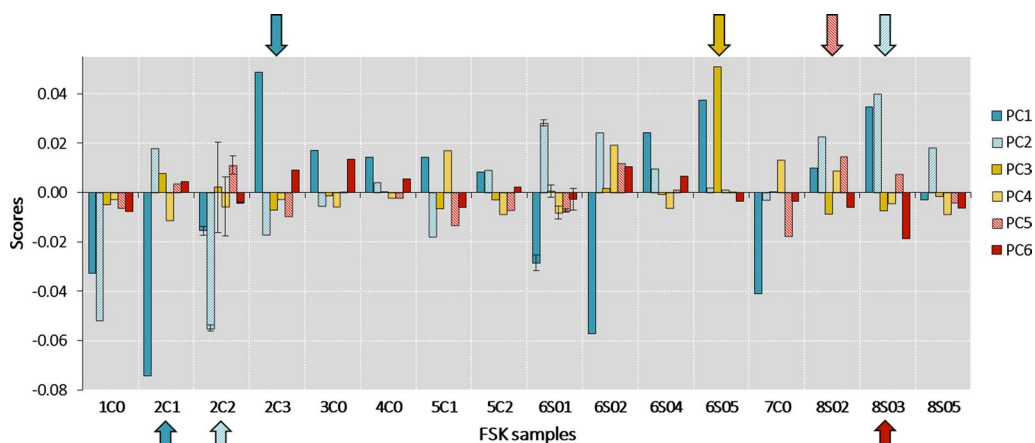


as endocrine disruptors, are persistent and bioaccumulative, and have been found ubiquitously in the aquatic environment [29]. The positive PC2 loading coefficients are mostly described by a few peaks (#39 diterpenes and #64.d tetralin derivatives) which, as in the positive PC1 loading, is indicative for a dominating natural chemical fingerprint (Fig. 4).

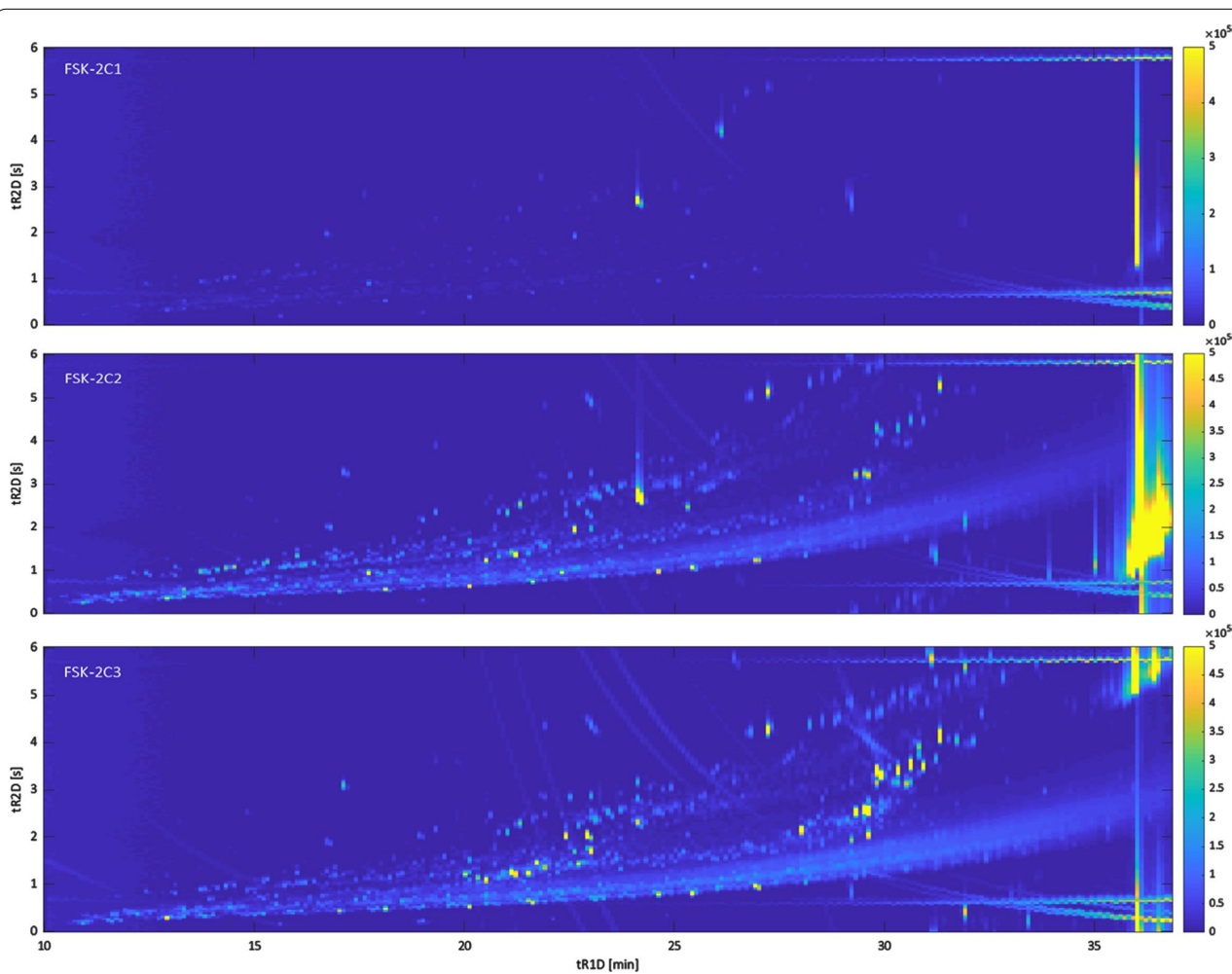
PC5 and PC6 explain merely 3.16% and 2.43% of the variation, respectively. They can still be useful to differentiate samples, e.g., with a higher relative abundance of lighter compounds, alkylphenols (#7), butylated hydroxytoluene (#19) and dibenzothiophene (#30) in the positive PC5 loading; and samples with a dominating range of peaks between 20 and 25 min in <sup>1</sup>D (Fig. 4). It seems that

specific types of diterpenes (#39) and tetralin derivatives (#64.d) can be distinguished in the positive and negative PC6 loading. However, examinations of the raw chromatograms revealed that peaks within the dotted white circle in PC5 and PC6 (Fig. 4) were misaligned in <sup>2</sup>D during data pre-processing. These misaligned regions therefore do not represent chemical variation of the data.

The score plot along the flow direction of the channel with all relevant PCs is shown in Fig. 5 with the highest scoring samples in each PC indicated by an arrow. No particular pattern along the flow direction of the channel could be recognised according to the scores. Error bars are the standard deviation of the sample preparation duplicates and demonstrate high repeatability of the



**Fig. 5** Bar plot (PC1, PC2, PC5 and PC6) of the FSK sampling site. Error bars were derived from the standard deviation of the sample preparation replicates. The arrows indicate the highest scoring sample in the corresponding PC



**Fig. 6** Base peak chromatograms of the fortress channel samples 2C1 (0–10 cm), 2C2 (10–20 cm) and 2C3 (20–30 cm)

sample preparation and an acceptable pre-processing as part of the WPCA modelling [21]. Single sampling spots highlight that there were variations within the same sampling grid, e.g., samples 6S01, 02, 04 and 05 range from  $-0.055$  to  $0.038$  in PC1 (Fig. 5). The loading of PC3 (positive direction) describes mainly sample 6S05 (Additional file 1: Figure S4). Composite samples, on the other hand, describe an average of five sampling spots per sampling grid, such as for sampling grids 3C0 and 7C0. The considerable variation within one sampling grid is significant with respect to the sampling strategy. Therefore, the chromatograms highlight the importance of an adequate sampling strategy, particularly for sediments and soils with their high chemical heterogeneity and poor mobility of nonpolar contaminants [30].

Six out of 18 samples were prioritised for compound identification according to their scoring in the particular loadings, namely 2C1, 2C2, 2C3, 6S0-5, 8S0-2 and 8S0-3. These samples build the corners or of the dataset in the particular loadings. Alkanes (#5), dichloro-diphenyl-dichloroethane (DDD, #24), polychlorinated biphenyls (PCBs, #51) and PAHs were more prevalent with increasing sampling depth. Dichloro-diphenyl-trichloroethane (DDT) and PCBs are POPs and have been phased out in Europe in the 1970s. Therefore, it comes to no surprise to find these pollutants at relatively higher abundance in lower sediment levels compared to the top-layer (<10 cm). Yet, this was difficult to recognise from the loading plots alone (Fig. 4). Other compounds in Table 1 are not less important but were found in several samples such as decalin and its derivatives (#26) which are industrial solvents used in fuel additives.

Sample 2C1 (0–10 cm) had a high relative abundance of alkylphenols (#7), alkylated benzenes (#10.4), non-alkylated styrene (#62) and tetralin (#64) (Table 1, cf. Fig. 6). Sample 2C2 was relatively different from the rest of the samples which is implied by its relatively large negative score (between  $-0.052$  and  $-0.055$ ) in PC2 compared to the rest of the samples (Fig. 5). The only diterpenes that could be identified at Level 3 were 10,18-bisnorabieta-8,11,13-triene ( $C_{18}H_{26}$ , #39.1) and Methyl-10,18-bisnorabieta-8,11,13-triene ( $C_{19}H_{28}$ , #39.2). The NIST library search suggested polycyclic musk (included in the term ‘tetralin derivatives’, #64.d) such as tonalide or versalide (used in personal care products), of which the former is associated with long-term adverse effects to the aquatic life. It was not possible, however, to confirm the tentative identification without the target analysis with standards.

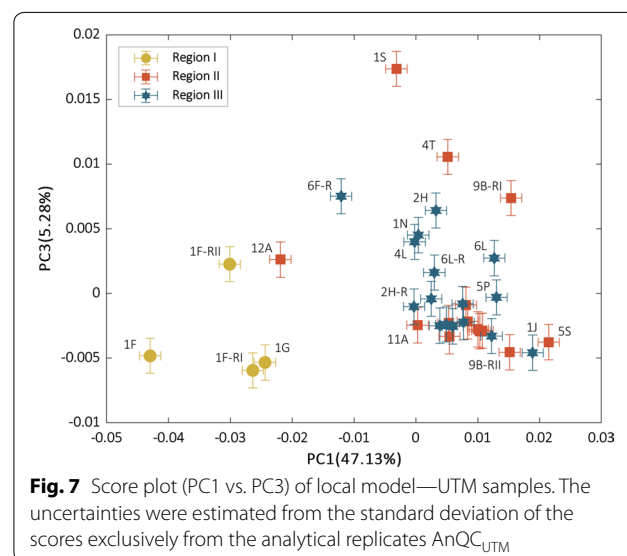
#### Local models—lake (UTM) site

The model for the lake site was built with four PCs and described 65.0% of the explained variance. The PC1 and

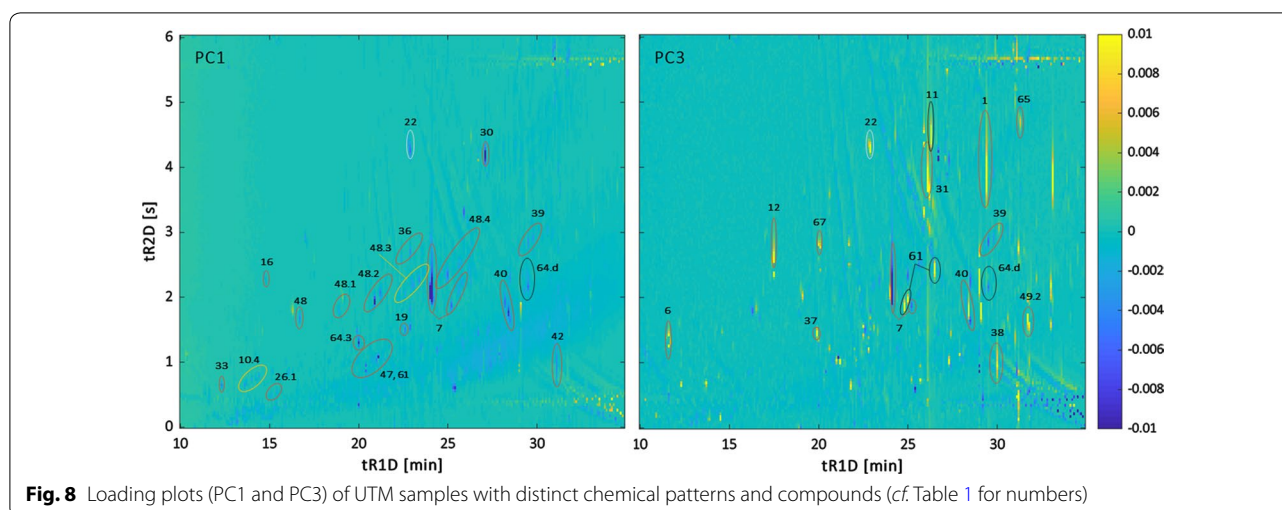
PC3 score and loading plots are shown in Figs. 7 and 8, respectively (Additional file 1: Figure S5 for the score plot PC2 vs. PC4). These PCs are the most descriptive PCs without retention time shifts as it was the case for PC2. Samples collected close to the channel outlet (samples 1F and 1G, including replicates, Fig. 1) and samples 12A and 6F-R, have a negative PC1 score and describe a very similar weathered oil fingerprint as samples 1C0 and 2C2 from the channel, which confirms the observations from PC2 from the global model. The GC  $\times$  GC–HRMS raw chromatograms of 1F and 12A confirmed this (not shown). Most of the samples from the centre of the lake (*Region III*) have very similar chemical fingerprints. That represents the averaged chemical fingerprint of all the samples from this site, contrary to the samples in the corners in Fig. 7. Samples close to a road (*Region II*) demonstrate a high variation in the score plot (Fig. 7) such as 1S and 4T (closest to the road on the east of the lake) and 12A and 9B-RI (from the south of the lake).

Despite the 1:5 dilution of the channel samples, compound prevalence in the lake were still considerably lower than compared to the fortress channel samples. The prominent peaks (e.g., #1, #11 and #31) in the PC3 loading (Fig. 8) indicate potential insufficient remobilisation in  $^2D$  of these compounds or that these compounds were tailing. Nevertheless, compound identification was possible.

Alkylphenols (#7) were highly abundant in samples close to the outlet of the fortress channel (*Region I*) as it can be seen by the negative PC1 and PC3 loadings (Fig. 8). Interestingly, this was not the case for samples from the east, which is indicative of dilution from the channel towards the lake (Fig. 1). Diterpenes (#39) were relatively



**Fig. 7** Score plot (PC1 vs. PC3) of local model—UTM samples. The uncertainties were estimated from the standard deviation of the scores exclusively from the analytical replicates AnQC<sub>UTM</sub>



more abundant in samples from *Region I*. Further, fatty acid methyl esters (#40) were detected in higher relative abundance in this region as were alkylated benzenes (#10.4), naphthalenes (#48–48.4) and many compounds that were also identified in the fortress channel.

Samples with positive PC3 scores (such as 1S, 4T and 9B-RII) contain relatively more of heterocyclic compounds, including 2-mercaptobenzothiazole (#1), benzoic acid esters (#11), benzothiazole (#12), dibenzylamine (#31), alkylated naphthothiophenes (#49.2), hydroxylated or sulphur-containing sesquiterpenes (#61) and vanillins (#67). Most of these compounds have been reported elsewhere in freshwater sediments [14, 31]. Many of these compounds are natural plant metabolites; others such as the alkylated naphthothiophenes are potentially derived from a petroleum source. Benzothiazole and 2-mercaptobenzothiazole are used in various industrial processes, such as for rubber vulcanisation or as a corrosion inhibitor. These compounds are biologically active and potential aquatic toxins [32]. Tire wear particles were identified as a potentially significant source in the environment [32]. Dibenzylamine (#31) is also an additive and by-product from the production of rubber. Benzothiazole, its derivatives and dibenzylamine were detected in sediment samples and associated with a rubber production factory in China [33]. The higher relative abundance of these compounds in samples 1S and 4T is most likely linked to the proximity to the highway, which is one of the four major routes to Copenhagen and among the ten busiest highways in Denmark [34]. The contamination source and, to a large extent, the chemical fingerprint in PC 3 loading (Fig. 8), could therefore be defined as traffic-related, perhaps even more specifically to tire wear

particles. The impact on tire wear particles in the aquatic environment was recently reviewed by Wagner et al. [35].

## Conclusions

Nontarget screening of urban freshwater sediment was performed by GC × GC–HRMS and pixel-based chemometric analysis. The study shows that pixel-based PCA on the 2D-BPCs without prior selection of specific ions can be a powerful tool for the NTS of sediments. The tiered NTS workflow included (i) a pre-screening of the sample extract raw chromatograms and taking a decision on the modelling strategy; (ii) a global pixel-based PCA model to obtain a map of the overall variation of the sampling area; (iii) local models of each sampling site for a more thorough investigation and identification; (iv) source identification and identification of prioritised peaks. A proper pre-processing of the data is crucial before building the models. It was possible to describe spatial and in-depth variation, and specific chemical fingerprints such as natural, weathered oil or high-traffic/tire wear. The prioritisation for tentative identification was not based only on peak intensity, but also on the highest variation between the samples. The pixel-based PCA prioritisation is primarily favouring samples (and thus, compounds) with a varying abundance and omits potential contaminants that are present in all the samples at a very similar concentration level. For the identification workflow, however, the prioritised sample was analysed as a whole which allowed the reporting of more compounds than visible in the PCA plots. The authors also would like to emphasise the importance of an appropriate sampling strategy regarding solid environmental samples due to

the considerable variation between different sampling spots as it was shown herein.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12302-020-00353-2>.

**Additional file 1.** Supporting information in relation to sampling, standards, data pre-processing, modelling and retention indice calculation.

**Additional file 2.** Detailed list of identified compounds in sediment samples from Copenhagen.

## Abbreviations

AnQC: Analytical quality control; BPC: Base peak chromatogram; DCM: Dichloromethane; EIC: Extracted ion chromatogram; EtOH: Ethanol; FSK: Fæstningskanalen (Fortress channel); GC–MS: Gas chromatography–mass spectrometry; GC × GC–HRMS: Two-dimensional gas chromatography–high-resolution mass spectrometry; MF: Match factor; MeOH: Methanol; NTS: Nontarget screening; PAH: Polycyclic aromatic hydrocarbon; PC: Principal component; PCA: Principal component analysis; PCB: Polychlorinated biphenyl; PLE: Pressurised liquid extraction; POPs: Persistent organic pollutants; QTOF: Quadrupole-time-of-flight; RSD: Relative standard deviation; UTM: Utterslev Mose (lake); WPCA: Weighted principal component analysis.

## Acknowledgements

The authors would like to express their gratitude to Esther Boll, Pablo Denti and Chiara Lucariello for their help during sampling and sample preparation. Giorgio Tomasi is acknowledged for his assistance in the data analysis.

## Authors' contributions

JHC contributed to the designing of the sampling campaign. JSL collected the samples and performed the sample extraction. GLA and JSL performed the chemical analysis with GC × GC–HRMS. GLA executed the pre-processing and pixel-based chemometric modelling in Matlab. JSL and GLA did the identification workflow of prioritised peaks and chemical interpretation. JHC significantly contributed to the model and data evaluation, and chemical interpretation. JSL wrote the draft. GLA and JHC critically revised the manuscript. All authors read and approved the final manuscript.

## Funding

We acknowledge financial support by Innovation Fund Denmark for the GANDALF project under Grant Agreement No. 5150-00008B.

## Availability of data and materials

The supporting information provides additional information about the standards and parameters in MassHunter Unknowns Analysis; additional raw chromatograms, loading plots for the global and local models, and the score plot for the local model of the lake. An excel sheet with all identified compounds is provided in Additional file 2.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 31 December 2019 Accepted: 27 April 2020

Published online: 11 May 2020

## References

- Connor R, Uhlenbrook S, Koncagül E, Ortigara ARC (2017) The United Nations world water development report 2017—wastewater, an untapped resource; executive summary. WWAP: United Nations
- Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, von Gunten U, Wehrli B (2006) The challenge of micropollutants in aquatic systems. *Science* 313:1072–1077. <https://doi.org/10.1126/science.1127291>
- Veenaas C, Bignert A, Liljelind P, Haglund P (2018) Nontarget screening and time-trend analysis of sewage sludge contaminants via two-dimensional gas chromatography–high resolution mass spectrometry. *Environ Sci Technol* 52:7813–7822. <https://doi.org/10.1021/acs.est.8b01126>
- European Parliament and Council (2013) Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy Text with EEA relevance. *Official Journal of the European Union* L226/1
- Reiner EJ, Jobst KJ, Megson D, Dorman FL, Focant J-F (2014) Analytical methodology of POPs. In: O'Sullivan G, Sandau C (eds) *Environmental forensics for persistent organic pollutants*. Elsevier, Amsterdam
- Hollender J, Schymanski EL, Singer HP, Ferguson PL (2017) Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ Sci Technol* 51:11505–11512. <https://doi.org/10.1021/acs.est.7b02184>
- Di Carro M, Magi E, Massa F, Castellano M, Mirasole C, Tanwar S, Olivari E, Povero P (2018) Untargeted approach for the evaluation of anthropic impact on the sheltered marine area of Portofino (Italy). *Mar Pollut Bull* 131:87–94. <https://doi.org/10.1016/j.marpolbul.2018.03.059>
- Muscalu AM, Gorecki T (2018) Comprehensive two-dimensional gas chromatography in environmental analysis. *Trends Anal Chem* 106:225–245. <https://doi.org/10.1016/j.trac.2018.07.001>
- Hashimoto S, Zushi Y, Takazawa Y, Ieda T, Fushimi A, Tanabe K, Shibata Y (2018) Selective and comprehensive analysis of organohalogen compounds by GC × GC–HRToFMS and MS/MS. *Environ Sci Pollut Res Int* 25:7135–7146. <https://doi.org/10.1007/s11356-015-5059-5>
- Ubukata M, Jobst KJ, Reiner EJ, Reichenbach SE, Tao Q, Hang J, Wu Z, Dane AJ, Cody RB (2015) Non-targeted analysis of electronics waste by comprehensive two-dimensional gas chromatography combined with high-resolution mass spectrometry: using accurate mass information and mass defect analysis to explore the data. *J Chromatogr A* 1395:152–159. <https://doi.org/10.1016/j.chroma.2015.03.050>
- Zushi Y, Hashimoto S, Tanabe K (2016) Nontarget approach for environmental monitoring by GC × GC–HRToFMS in the Tokyo Bay basin. *Chemosphere* 156:398–406. <https://doi.org/10.1016/j.chemosphere.2016.04.131>
- Blum KM, Gallampois C, Andersson PL, Renman G, Renman A, Haglund P (2019) Comprehensive assessment of organic contaminant removal from on-site sewage treatment facility effluent by char-fortified filter beds. *J Hazard Mater* 361:111–122. <https://doi.org/10.1016/j.jhazmat.2018.08.009>
- Ortiz-Almirall X, Pena-Abaurrea M, Jobst KJ, Reiner EJ (2016) Chapter 14—nontargeted analysis of persistent organic pollutants by mass spectrometry and GC × GC. In: Pérez S, Eichhorn P, Barceló D (eds) *Comprehensive analytical chemistry*. Elsevier, Amsterdam
- Bastos PM, Haglund P (2012) The use of comprehensive two-dimensional gas chromatography and structure–activity modeling for screening and preliminary risk assessment of organic contaminants in soil, sediment, and surface water. *J Soils Sediments* 12:1079–1088. <https://doi.org/10.1007/s11368-012-0533-x>
- Schymanski EL, Singer HP, Slobodnik J, Ipolyi IM, Oswald P, Krauss M, Schulze T, Haglund P, Letzel T, Grosse S et al (2015) Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal Bioanal Chem* 407:6237–6255. <https://doi.org/10.1007/s00216-015-8681-7>
- Gago-Ferrero P, Schymanski EL, Hollender J, Thomaidis NS (2016) Chapter 13—nontarget analysis of environmental samples based on liquid

- chromatography coupled to high resolution mass spectrometry (LC-HRMS). In: Pérez S, Eichhorn P, Damià B (eds) *Comprehensive analytical chemistry*. Elsevier, Amsterdam
17. Chiaia-Hernandez AC, Gunthardt BF, Frey MP, Hollender J (2017) Unraveling contaminants in the anthropocene using statistical analysis of liquid chromatography-high-resolution mass spectrometry nontarget screening data recorded in lake sediments. *Environ Sci Technol* 51(21):12547–12556. <https://doi.org/10.1021/acs.est.7b03357>
  18. Pierce KM, Parsons BA, Synovec RE (2015) Chapter 10—pixel-level data analysis methods for comprehensive two-dimensional chromatography. In: de la Peña AM, Goicoechea HC, Escandar GM, Olivieri AC (eds) *Data handling in science and technology*. Elsevier, Amsterdam
  19. Alexandrino GL, Malmborg J, Augusto F, Christensen JH (2019) Investigating weathering in light diesel oils using comprehensive two-dimensional gas chromatography-high resolution mass spectrometry and pixel-based analysis: possibilities and limitations. *J Chromatogr A* 1591:155–161. <https://doi.org/10.1016/j.chroma.2019.01.042>
  20. Zhang D, Huang X, Regnier FE, Zhang M (2008) Two-dimensional correlation optimized warping algorithm for aligning GC × GC-MS data. *Anal Chem* 80:2664–2671. <https://doi.org/10.1021/ac7024317>
  21. Christensen JH, Tomasi G (2007) Practical aspects of chemometrics for oil spill fingerprinting. *J Chromatogr A* 1169:1–22. <https://doi.org/10.1016/j.chroma.2007.08.077>
  22. Christensen JH, Hansen AB, Tomasi G, Mortensen J, Andersen O (2004) Integrated methodology for forensic oil spill identification. *Environ Sci Technol* 38:2912–2918. <https://doi.org/10.1021/es035261y>
  23. 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. *Environ Sci Technol* 48:2097–2098. <https://doi.org/10.1021/es5002105>
  24. Rowdhwal SSS, Chen J (2018) Toxic effects of di-2-ethylhexyl phthalate: an overview. *Biomed Res Int* 2018:1750368. <https://doi.org/10.1155/2018/1750368>
  25. Introduction to dictionary of natural products (DNP) (1997) In: Buckingham J (ed) *Dictionary of natural products v28.1*. CRC Press, Taylor & Francis Group, Boca Raton
  26. Seiler TB, Schulze T, Hollert H (2008) The risk of altering soil and sediment samples upon extract preparation for analytical and bio-analytical investigations—a review. *Anal Bioanal Chem* 390:1975–1985. <https://doi.org/10.1007/s00216-008-1933-z>
  27. Pollo BJ, Alexandrino GL, Augusto F, Hantao LW (2018) The impact of comprehensive two-dimensional gas chromatography on oil & gas analysis: recent advances and applications in petroleum industry. *Trends Anal Chem* 105:202–217. <https://doi.org/10.1016/j.trac.2018.05.007>
  28. Gallotta FD, Christensen JH (2012) Source identification of petroleum hydrocarbons in soil and sediments from Iguacu River Watershed, Parana, Brazil using the CHEMSIC method (CHEMometric analysis of Selected Ion Chromatograms). *J Chromatogr A* 1235:149–158. <https://doi.org/10.1016/j.chroma.2012.02.041>
  29. Priac A, Morin-Crini N, Druart C, Gavaille S, Bradu C, Lagarrigue C, Torri G, Winterton P, Crini G (2017) Alkylphenol and alkylphenol polyethoxylates in water and wastewater: a review of options for their elimination. *Arab J Chem* 10:S3749–S3773. <https://doi.org/10.1016/j.arabj.2014.05.011>
  30. Danish Standards Foundation (2013) Representative sampling—horizontal standard (DS3077:2013). 2nd edition, Copenhagen, Denmark, 26 Aug 2013
  31. Schlabach M, Haglund P, Reid MJ, Rostkowski P (2017) Suspect screening in nordic countries—point sources in city areas. In: TemaNord. Nordic Council of Ministries, Copenhagen. <https://doi.org/10.6027/tn2017-561>
  32. Avagyan R, Sadiqsis I, Bergvall C, Westerholm R (2014) Tire tread wear particles in ambient air—a previously unknown source of human exposure to the biocide 2-mercaptobenzothiazole. *Environ Sci Pollut Res Int* 21:11580–11586. <https://doi.org/10.1007/s11356-014-3131-1>
  33. Xiao H, Krauss M, Floehr T, Yan Y, Bahlmann A, Eichbaum K, Brinkmann M, Zhang X, Yuan X, Brack W et al (2016) Effect-directed analysis of aryl hydrocarbon receptor agonists in sediments from the three gorges reservoir, China. *Environ Sci Technol* 50:11319–11328. <https://doi.org/10.1021/acs.est.6b03231>
  34. Brinch-Pedersen C. (2019) Vi bliver flere og flere på vejene. Vejdirektoratet, Denmark. <https://www.vejdirektoratet.dk/side/traffikkens-udvikling-i-tal>. Accessed 08 Nov 2019
  35. Wagner S, Huffer T, Klockner P, Wehrhahn M, Hofmann T, Reemtsma T (2018) Tire wear particles in the aquatic environment—a review on generation, analysis, occurrence, fate and effects. *Water Res* 139:83–100. <https://doi.org/10.1016/j.watres.2018.03.051>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---