



UvA-DARE (Digital Academic Repository)

Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS)

Brock, O.; Helmus, R.; Kalbitz, K.; Jansen, B.

DOI

[10.1111/ejss.12894](https://doi.org/10.1111/ejss.12894)

Publication date

2020

Document Version

Final published version

Published in

European Journal of Soil Science

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Brock, O., Helmus, R., Kalbitz, K., & Jansen, B. (2020). Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS). *European Journal of Soil Science*, 71(3), 420-432. <https://doi.org/10.1111/ejss.12894>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

ORIGINAL ARTICLE

Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS)

Olaf Brock¹  | Rick Helmus¹ | Karsten Kalbitz² | Boris Jansen¹

¹Department of Ecosystem and Landscape Dynamics, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, the Netherlands

²Soil Resources and Land Use, Institute of Soil Science and Site Ecology, Technische Universität Dresden, Tharandt, Germany

Correspondence

Olaf Brock, Department of Ecosystem and Landscape Dynamics, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94240, Sciencepark 904, 1090 GE Amsterdam, the Netherlands.

Email: o.p.brock@uva.nl

Funding information

Stichting voor de Technische Wetenschappen, Grant/Award Number: 13883

Abstract

Leaf litter-derived dissolved organic matter (DOM) is an important source of organic matter entering the mineral soil, but characterization of leaf litter DOM is often not detailed enough to understand DOM dynamics and processes at the molecular level. (Ultra-) high-resolution techniques such as Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), Orbitrap mass spectrometry and quadrupole time-of-flight (QTOF) mass spectrometry have brought a new level of detail to (D)OM research by providing molecular formulae and information on molecular structures. We present a novel approach for leaf litter DOM characterization that includes non-target screening using liquid chromatography (LC) coupled to high-resolution QTOF-MS. A method validation test showed that out of 26 spiked naturally occurring DOM molecules in a leaf litter DOM sample, 23 were found as features. After implementation of additional filtering to exclude certain combinations of heteroatoms (that are difficult to resolve even by the newest high-resolution MS instruments), including molecular formulae with P atoms, 22 molecular formulae were correctly assigned. Although a large part of the annotated molecular formulae was unique for the respective litter sources, the molecular compound class distribution was similar for deciduous, coniferous and mixed leaf litter DOM. Most intense masses were detected in the 100–300 Da mass range, showing a bias of QTOF-MS towards lower masses compared to FT-ICR-MS. The use of LC in combination with QTOF-MS for leaf litter (D)OM characterization provides, together with Orbitrap-MS, a more widely available and lower cost high-resolution MS alternative to FT-ICR-MS. Novel approach to the characterization and analysis of leaf litter derived DOM. Non-target screening using LC-QTOF-MS and the patRoon R package. Similar molecular compound class distribution for three leaf litter sources. High-resolution DOM characterization alternative to FT-ICR-MS.

Highlights

- Novel approach to the characterization and analysis of leaf litter-derived DOM.
- Non-target screening using LC-QTOF-MS and the patRoon R package.
- Similar molecular compound class distribution for three leaf litter sources.
- High-resolution DOM characterization alternative to FT-ICR-MS.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. European Journal of Soil Science published by John Wiley & Sons Ltd on behalf of British Society of Soil Science.

KEYWORDS

DOM, molecular compound classes, molecular formulae, organic matter characterization, Suwannee River fulvic acid

1 | INTRODUCTION

Although it is well known that the quantity of organic matter (OM) affects many soil functions (e.g., water storage and structural soil stability) the effects of OM composition on the functionality of the soil are much less known and the subject of scientific debate (Lehmann & Kleber, 2015; Schmidt et al., 2011). Because a significant and very reactive part of OM is in a dissolved form, all the above also applies to dissolved organic matter (DOM). Although DOM has been studied intensively over the past decades (Nebbioso & Piccolo, 2013; Rathgeb, Causon, Krachler, & Hann, 2017), determining key molecular characteristics of a given DOM mixture at the appropriate resolution remains an important challenge. For instance, the conventional description of the OM composition in functional groups such as alkyl-carbon (alkyl-C) and aromatic-C (Scheel, Haumaier, Ellerbrock, Rühlmann, & Kalbitz, 2008) is not detailed enough to understand the interaction of metal cations with DOM at the molecular level.

Nebbioso and Piccolo (2013) reviewed techniques that partially characterize (D)OM at the molecular level. Nuclear magnetic resonance (NMR) spectroscopy provides information on functional groups present in bulk natural OM (Scheel et al., 2008), but fails to provide molecular formulae and structures (Sleighter & Hatcher, 2007). Pyrolysis gas chromatography–mass spectrometry (pyr-GC–MS), liquid chromatography–mass spectrometry (LC–MS) and tetramethylammonium hydroxide (TMAH) thermochemolysis all fail to fully resolve the complexity of DOM mixtures due to relatively low instrumental resolutions and/or the formation of new products during sample preparation (Sleighter & Hatcher, 2007). (Ultra-) high-resolution techniques such as Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) (Cory et al., 2010; Hawkes, Dittmar, Patriarca, Tranvik, & Bergquist, 2016; Kujawinski & Behn, 2006), Orbitrap mass spectrometry (Hawkes et al., 2016; Patriarca, Bergquist, Sjöberg, Tranvik, & Hawkes, 2018; Petras et al., 2017; Remucal, Cory, Sander, & McNeill, 2012; Verkh, Rozman, & Petrovic, 2018) and quadrupole time-of-flight mass spectrometry (QTOF-MS) (Rathgeb et al., 2017; These, Winkler, Thomas, & Reemtsma, 2004) provide a much more detailed characterization of DOM. For QTOF-MS (and Orbitrap-MS), often coupled to a liquid chromatography (LC) system, this is especially the case when tandem mass spectrometry (MS/MS) is used to obtain fragment information. Coupling of LC to QTOF-MS is needed

to obtain simple enough mass spectra that provide a sufficiently high accuracy compared to instruments with a higher instrumental resolution such as FT-ICR-MS. The increased analytical power of both QTOF-MS and Orbitrap-MS (resolutions up to 50,000) has made these techniques potential alternatives to the less widely available, more costly, but analytically superior FT-ICR-MS (resolution in excess of 100,000) (Hawkes et al., 2016).

Non-target screening using high-resolution QTOF-MS is often used in the field of wastewater research (Verkh et al., 2018), but has not been tested yet for leaf litter or soil-derived DOM. We are aware of only a handful of studies that used LC-QTOF-MS to study OM. Schmitt-Kopplin et al. (2010) analysed extraterrestrial OM with a combination of FT-ICR-MS, NMR and ultra-performance (UP) LC-QTOF-MS (Schmitt-Kopplin et al., 2010). These et al. (2004) studied low-molecular-weight fulvic acids of different origin with size exclusion chromatography (SEC)-QTOF-MS. Rathgeb et al. (2017) used LC-QTOF-MS for non-target analysis of riverine DOM. Beside the different field of research – soil science instead of marine chemistry – our study differs from the study of Rathgeb et al. (2017) by having a largely automated approach to data analysis. We use the in-house-developed patRoon R-package (Helmus, 2018) to harmonize various open source and vendor data processing tools. The patRoon R package is open source and instrument independent.

We are unaware of studies in which liquid chromatography coupled to high-resolution MS (QTOF/Orbitrap) was used to characterize leaf litter-derived DOM at the molecular level. Leaf litter-derived DOM is an important source of OM entering the mineral soil and is therefore relevant for DOM dynamics in soils (Kaiser & Kalbitz, 2012) (e.g., in the pedogenesis of podzols) and represents an important fraction of OM input into rivers and lakes. QTOF-MS provides the possibility to obtain structural information at the molecular level by measuring mass fragments in the MS/MS mode and to combine this with insights from bulk information as visualized in Van Krevelen and related plotting techniques (Koch, Witt, Engbrodt, Dittmar, & Kattner, 2005). However, analysis of high-resolution MS data has brought new challenges regarding formula determination and data display based on the vast amount of information acquired during a given analysis (Tziotis, Hertkorn, & Schmitt-Kopplin, 2011).

Koch, Dittmar, Witt, and Kattner (2007) summarized three main challenges regarding formula determination when analysing high-resolution MS data. First, often more than one candidate formula can be calculated for each detected mass. Second, manual formula assignment for thousands of features – potential compounds based on retention and mass information – is extremely time consuming, but automated methods lack parameters for unambiguous formula determination. Third, independent validation with an external standard is impossible, because a complex molecular mixture with known composition that resembles natural organic matter is not available.

Reemtsma (2009) reviewed status and needs regarding molecular formula determination from high-resolution MS data (Reemtsma, 2009). A distinction is often made between two different strategies: (a) the brute force approach and (b) the formula extension/building block approach. In the brute force approach a list of possible chemical elemental formulae is calculated for each nominal mass. Observed masses are then compared to the masses of this list, and the closest fit is chosen as the most likely elemental formula. In the formula extension approach, formulae are first determined for a small subset of low-molecular-weight compounds, after which chemical and structural relationships among compounds are used to find formulae for higher masses. Examples of the latter strategy are the chemical building block approach (Koch et al., 2007) and Kendrick mass analysis (Kujawinski & Behn, 2006). To speed up formula assignment, Kujawinski and Behn (2006) automated parts of this process with the compound identification algorithm (CIA). Multiple computational tools have been developed to visualize (ultra-) high-resolution MS data. The Van Krevelen diagram is used frequently and specifies different compound classes based on their elemental hydrogen/carbon (H/C) and oxygen/carbon (O/C) ratios (Kim, Kramer, & Hatcher, 2003). Another way to plot high-resolution MS data is the carbon versus mass diagram (Reemtsma, 2010). Often the goal is to obtain structural information from the calculated formulae (e.g., by calculating the double bond equivalence (DBE)) (Koch & Dittmar, 2006).

With (ultra-) high-resolution techniques becoming more widely available in soil laboratories, our aim was to develop and test a high-resolution MS alternative method to FT-ICR-MS for the characterization of DOM extracted from leaf litter using non-target screening by liquid chromatography (LC) coupled to high-resolution QTOF-MS. We describe a workflow from data (pre-) processing to data visualization/reporting and show results for a set of leaf litter samples from different vegetation sources. The non-target screening analysis software patRoon (Helmus, 2018) was tested and validated by spiking a mixed leaf litter DOM sample with 26 known compounds that occur in natural DOM.

2 | MATERIALS AND METHODS

2.1 | Sample selection and used chemicals

Leaf litter samples were collected from three podzol sites in the provinces Brabant and Limburg in the Netherlands (Table S1 in File S1). Podzols were chosen because leaf litter-derived DOM plays an important role in the pedogenesis of podzols. The three locations were selected for their dominant vegetation source: deciduous trees, coniferous trees and a mix of both tree types. In addition, we characterized the Suwannee River Fulvic Acid (SRFA) Standard II (2S101F) (International Humic Substances Society, 2017) as a well-documented standard to represent a complex DOM mixture, to allow comparison of our results with other high-resolution MS studies that also included the SRFA standard. We used ultrapure ULC/MS grade methanol (Biosolve, Valkenswaard, the Netherlands), acetic acid (Merck KGaA, Darmstadt, Germany), sodium hydroxide (Sigma-Aldrich, Zwijndrecht, the Netherlands) and ultrapure water (MK2-Analytic, ELGA, Veolia, Lane End, UK).

2.2 | Sample preparation

Leaf litter samples were dried at 40°C and shredded. High-speed centrifuge tubes were filled with 2.5 g of the shredded material. Three replicates were prepared per sample. Water extracts were made by adding 25 mL ultrapure water to the shredded material in a 1:10 weight-to-volume ratio. Four procedural blanks only contained ultrapure water. All tubes were shaken first manually and then overnight on a shaking plate. All samples were centrifuged at high speed (6,800 × g) for 20 min and then filtered to <0.2 µm (ME24 mixed cellulose ester membrane; Whatman, Little Chalfont, Buckinghamshire, UK). Filtrates were stored at 4°C. Samples were diluted 10 times to obtain an optimal MS detection range and prevent overloading of the MS detector. SRFA samples were prepared by dissolving freeze-dried solid product in ultrapure water and were measured directly with LC-QTOF-MS. Concentration effects were tested by comparing four different SRFA concentrations: 0.5, 1.0, 2.5 and 5.0 mg ml⁻¹.

2.3 | Instrumental settings

Analyses were conducted with an ultra-high-performance liquid chromatography (UHPLC) system (Nexera, Shimadzu, Den Bosch, the Netherlands) coupled to a high-resolution QTOF-MS (maXis 4G upgraded with an HD collision cell, Bruker Daltonics, Leiderdorp, the Netherlands). Analytes were separated using an Acquity UHPLC CSHC18 column (150 mm × 2.1 mm; 1.7 µm; Waters, Etten-Leur, the Netherlands). The injection volume was 20 µL and the flow rate was 0.2 mL minute⁻¹. The mobile phase consisted

of (A) ultrapure water with 0.05% acetic acid and (B) methanol. At the start, elution was performed with 5% of mobile phase B. In 15 min the percentage of mobile phase B was gradually increased to 100%. After 10 min of B at 100%, the mobile phase was reset to 5% in 1 min. The column was equilibrated for 8 min prior to every injection to re-establish initial mobile phase conditions. The column oven temperature was 40°C and the auto sample tray temperature was 15°C. After separation by UHPLC, eluting compounds were ionized by an electrospray ionization (ESI) source operating in negative mode. The MS detector was calibrated before starting an analysis batch and additionally prior to each injection. This was achieved by infusing a 50- μ M sodium acetate solution in water–methanol (1:1, v/v) directly into the source. In addition, automatic introduction of the calibration solution at the start of each run allowed post-recalibration of the MS data. Approximately the first 2 min of the chromatographic run (the part after the calibration and before the LC dead volume) were sent to the waste,

using a switch valve that connects the LC and MS systems, to avoid contamination by introducing unretained compounds into the mass spectrometer. Our instrumentation had a resolving power of 30,000 to 60,000 and full width at half maximum (FWHM) was 0.03 m/z. MS/MS data were recorded in data-dependent (Auto) MS/MS mode and used in formula determination. Data were scanned over a mass range of 80 to 1,200 m/z with a scan rate of 4 Hz. The focus on low-molecular-weight compounds is justified because high-molecular-weight compounds may fragment in low-molecular-weight subunits (These et al., 2004) and a recent study showed that at least 60% of DOM molecules are <1 kDa (Petras et al., 2017) (see part B for more details on the instrumental MS settings).

2.4 | Non-target screening

The workflow (Figure 1) for (semi-) automated non-target screening was developed in R (R Core Team, 2017) and

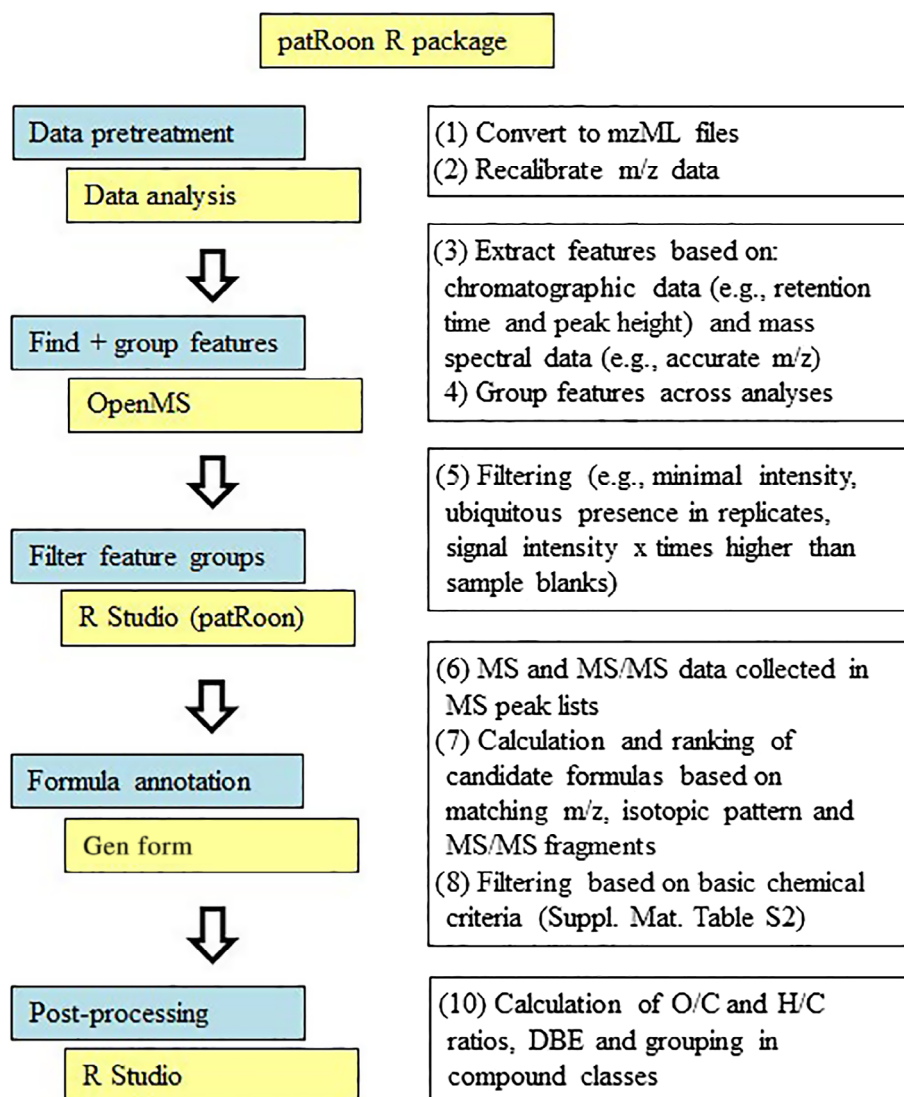


FIGURE 1 A workflow is shown of the developed non-target screening method. Steps are explained in the Materials and Methods section. DBE, double bond equivalence; MS, mass spectrometry; H/C, hydrogen/carbon ratio; O/C, oxygen/carbon ratio

consisted of the following steps: data pretreatment (automatic *m/z* recalibration to avoid mass drift and export to *mzML* files), extraction and alignment of chromatographic and mass spectral data, rule-based filtering based on minimal intensity, absence in sample blanks and ubiquitous presence in replicates, and automatic calculation of molecular formulae based on high-resolution MS and MS/MS data. Data were processed with DataAnalysis 4.4 (Bruker Daltonics, Bremen, Germany), GenForm (Meringer, Reinker, Zhang, & Muller, 2011) and OpenMS (Röst et al., 2016) software. These open source and vendor data processing tools for non-target analysis were harmonized by the in-house-developed (open source and instrument independent) patRoon R package (Helmus, 2018).

Features were extracted from the pretreated data by automatic identification of chromatographic peaks using the OpenMS function “FeatureFinderMetabo”. OpenMS is an open-source software library for LC/MS data management and analysis. Although other feature finder algorithms (e.g., XCMS) exist, pilot tests (results not shown) revealed that for our data the OpenMS feature finder resulted in the highest number of features and was most efficient computational wise. The most important settings for feature finding were a noise intensity threshold (“noise_threshold_int”) of 1,000, a maximum chromatographic peak length (“max_trace_length”) of 240 s and an allowed mass deviation of 3 ppm. A balance between quantity and quality was visually determined by applying settings that combined a high total number of features with a low number of features assigned to split peaks or noise (examples in Figure S8 in File S1). These settings were validated and tested by measuring spiked DOM samples (see method validation test). Equal features found across samples were grouped (Figure 1). Those “feature groups” were then filtered on peak intensity (minimum 5,000), ubiquitous presence in replicates and a signal intensity at least five times higher than sample blanks (Figure 1).

Molecular formulae were calculated automatically for all remaining feature groups. For this, first, MS peak lists were generated from the feature groups with the “generateMSpeakLists” function from the patRoon R package (Helmus, 2018). This function uses the *mzR* package to obtain MS and MS/MS spectra and averages and filters the data. Candidate formulae were calculated with the “generateFormulas” function (patRoon), which uses GenForm, an open source implementation of molecular formula calculation by high-resolution MS and MS/MS data (Meringer et al., 2011). GenForm uses the accurate mass of the feature groups to calculate candidate formulae and scores them based on matched theoretical/measured isotopic patterns and, if available, explained MS/MS fragments (Figure 1). We selected carbon, hydrogen, nitrogen, oxygen,

phosphorus and sulphur (CHNOPS) as the elements to be included in formula calculation. However, following Hawkes et al. (2016) we did not consider certain arrangements of heteroatoms (Table S2 in File S1), as they are difficult to resolve even by the newest HRMS instruments (Hawkes et al., 2016). Candidate formulae were also excluded when they did not meet basic chemical criteria (Table S2 in File S1) previously defined by Kujawinski and Behn (2006) and Koch & Dittmar (2006). For these filter steps it was necessary to calculate first the atomic ratios (O/C, H/C, N/C, S/C and P/C) and the double bond equivalence (DBE) (Koch & Dittmar, 2006). Feature groups without assigned candidate formulae were removed prior to further data processing. For each individual feature group, the remaining candidate formulae were evaluated and the highest ranked candidate formula was selected for post-processing and data visualization.

2.5 | Post-processing and data visualization

Assigned molecular formulae were grouped into different molecular compound classes based on previously published definitions (Abdulla, Sleighter, & Hatcher, 2013) (Table S3 in File S1). We also calculated the percentages of assigned molecular formulae that contained at least one heteroatom (N, S and/or P). Separate subsets of the total input data were created for each unique sample group, which included the replicates. Data were not scaled to the average peak intensities, because with ESI-MS, comparing intensities between different chemical compounds does not reliably indicate their relative abundances. The median was used to describe the elemental composition (CHNOPS), elemental ratios (O/C and H/C), molecular weights and DBE based on the assigned molecular formulae for each sample.

2.6 | Method validation test

To test the patRoon software (Helmus, 2018), 26 compounds known to occur in natural DOM (Table S5 in File S1) were spiked in a DOM sample extracted from the mixed leaf litter. The spiked leaf litter DOM samples were measured with LC-QTOF-MS in negative ionization mode using the settings described earlier. The above-described workflow was applied to process the data and retrieve the spiked compounds as features (peaks) in a realistic natural DOM background signal. Furthermore, molecular formula determination was tested by comparing the assigned molecular formulae with the target formulae. The spiked solution of the method validation test was also used to optimize and test the feature-finding parameters (results not shown) by a design-of-experiment methodology based on the IPO R package (Libiseller et al., 2015).

3 | RESULTS

3.1 | Method validation test

The method validation test showed that 23 of the 26 spiked compounds were detected as features. Malic acid and gallic acid could not be detected, which may be due to relatively high detection limits of these compounds (Figure 2 and Table S5 in File S1). Molecular formula calculation with patRoön was successful for 17 of the 23 identified compounds. For the other six compounds the first candidate formula was wrong. However, for five of these six compounds

the second candidate formula (ranked on scores), which did not contain P, was the correct formula (Table S5 in File S1).

3.2 | Suwannee River fulvic acid

Most molecular characteristics of the SRFA standard (Table 1) were similar over the range of concentrations used (0.5–1.0–2.5–5.0 mg ml⁻¹), although the number of assigned formulae increased from 151 for the lowest concentration to 476 for the highest concentration (Table 1). The O/C ratio increased from 0.67 to 0.82 and the H/C ratio

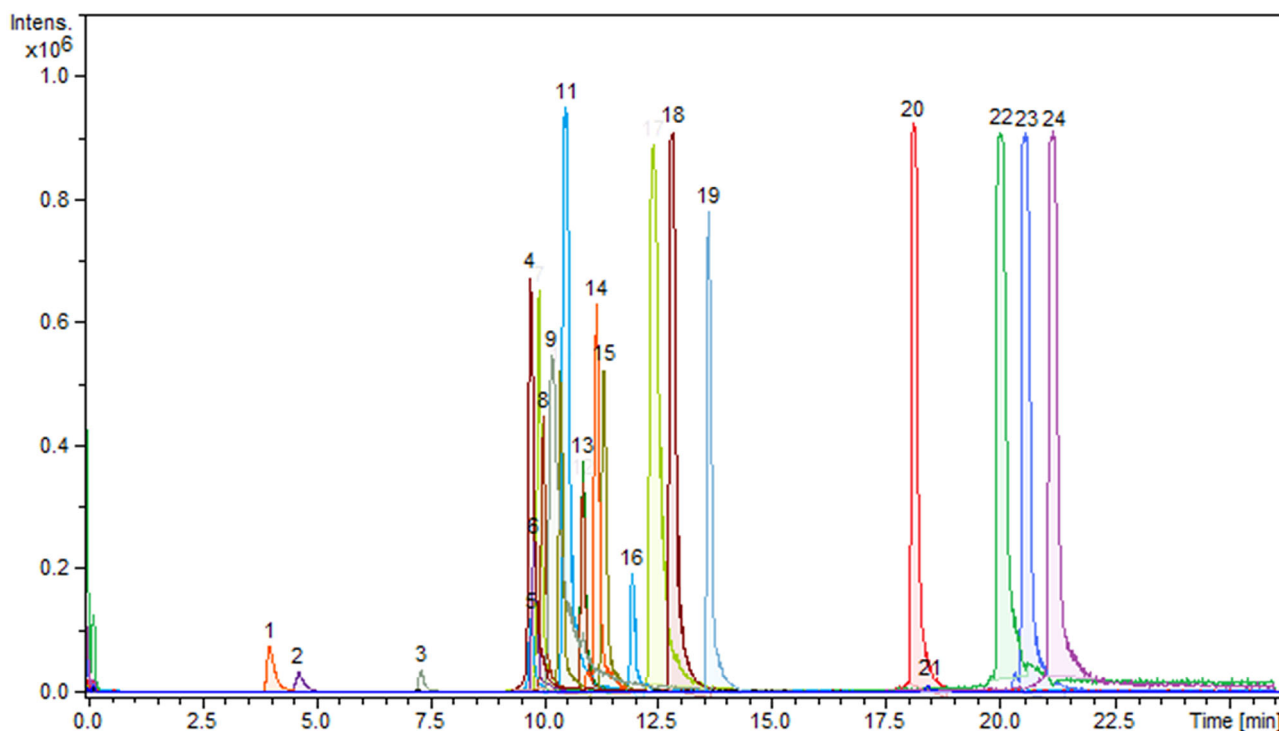


FIGURE 2 Extracted ion chromatogram of the 23 spiked compounds. Compound identification is provided in Table S5 in File S1

TABLE 1 Mean organic matter composition of the Suwannee River fulvic acid and leaf litter samples

Sample	MF ^a (<i>n</i>)	C# ^b	H# ^b	O# ^b	O/C ^c	H/C ^c	DBE ^d	MW ^f /Da
SRFA ^e (0.5 mg/mL)	151	7 (0.35)	8 (0.67)	4 (0.17)	0.67 (0.02)	1.33 (0.04)	3 (0.22)	176 (6)
SRFA (1.0 mg/mL)	204	7 (0.29)	8 (0.50)	5 (0.16)	0.75 (0.02)	1.33 (0.04)	3 (0.22)	190 (5)
SRFA (2.5 mg/mL)	283	7 (0.22)	8 (0.38)	5 (0.15)	0.80 (0.01)	1.20 (0.03)	4 (0.17)	186 (4)
SRFA (5.0 mg/mL)	476	8 (0.16)	8 (0.26)	6 (0.12)	0.82 (0.01)	1.00 (0.02)	5 (0.14)	202 (3)
Litter (deciduous)	684	12 (0.25)	16 (0.36)	6 (0.17)	0.55 (0.01)	1.43 (0.02)	4 (0.16)	261 (6)
Litter (coniferous)	949	13 (0.22)	18 (0.33)	5 (0.12)	0.43 (0.01)	1.43 (0.01)	5 (0.13)	266 (4)
Litter (mix)	667	12 (0.25)	17 (0.38)	5 (0.11)	0.43 (0.01)	1.44 (0.02)	5 (0.15)	254 (5)

^aNumber (*n*) of annotated molecular formulae.

^bMean numbers of carbon, hydrogen and oxygen numbers. Standard errors of the mean are shown in brackets.

^cAtomic ratios of oxygen and hydrogen to carbon.

^dDouble bond equivalence (Koch & Dittmar, 2006).

^eSuwannee River fulvic acid.

^fMolecular weight in Da (median).

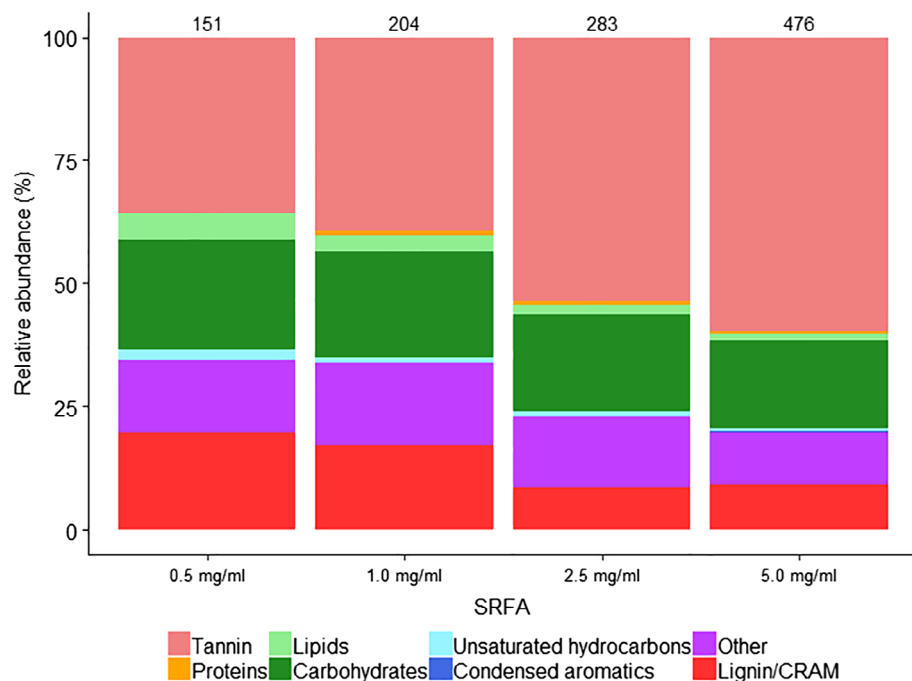


FIGURE 3 Compound class distribution for the Suwannee River fulvic acid standard. For the molecular compound class definitions see Table S3 in File S1. CRAM, carboxyl-rich alicyclic molecules; SRFA, Suwannee River fulvic acid

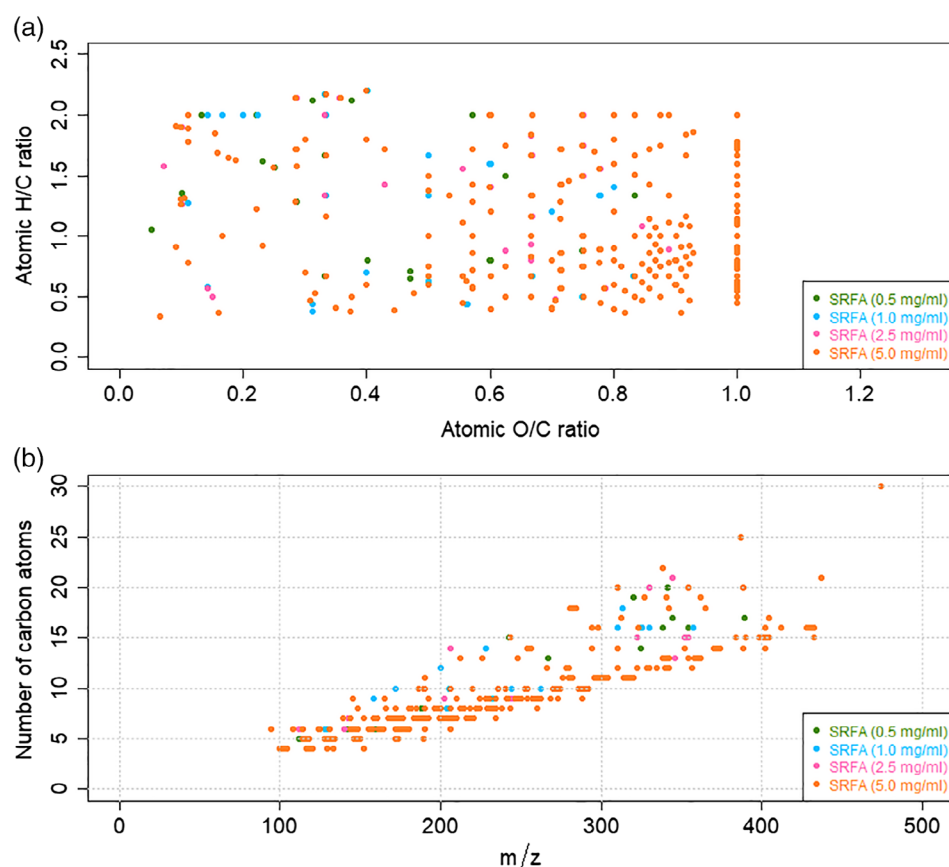


FIGURE 4 Van Krevelen plot (top) and carbon versus mass plot (bottom) for the SRFA standard. Each dot represents an annotated molecular formula. Formulae found for multiple samples are overlapping. H/C, hydrogen/carbon ratio; O/C, oxygen/carbon ratio; SRFA, Suwannee River fulvic acid

decreased from 1.33 to 1.00 from low to high SRFA concentration (Table 1). The increase in median DBE from 3 to 5 (Table 1) coincided with an increase in the relative abundance of tannin compounds (Figure 3), which are characterized by relatively high DBE values. All assigned molecular

formulae were plotted in a Van Krevelen diagram and a carbon versus mass diagram (Figure 4). In the Van Krevelen diagram 76.7% of all assigned formulae had O/C values >0.6, which is the region that includes tannin-derived compounds (Figure 4). The carbon versus mass diagram showed

FIGURE 5 Compound class distribution for the leaf litter samples. For the molecular compound class definitions see Table S3 in File S1. CRAM, carboxyl-rich alicyclic molecules

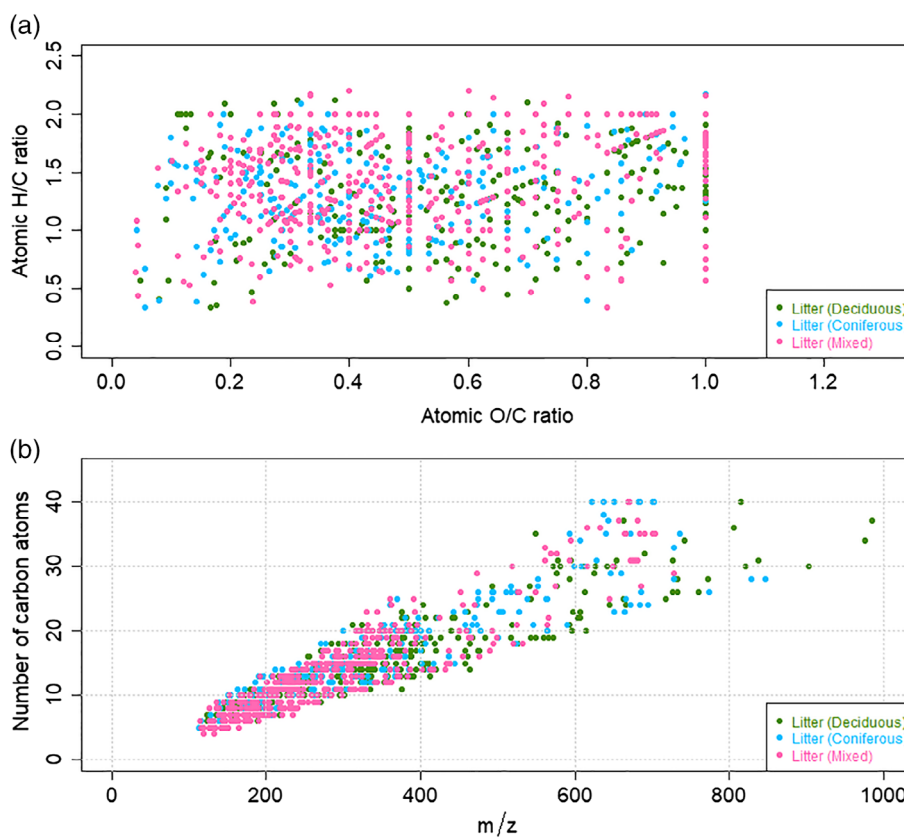
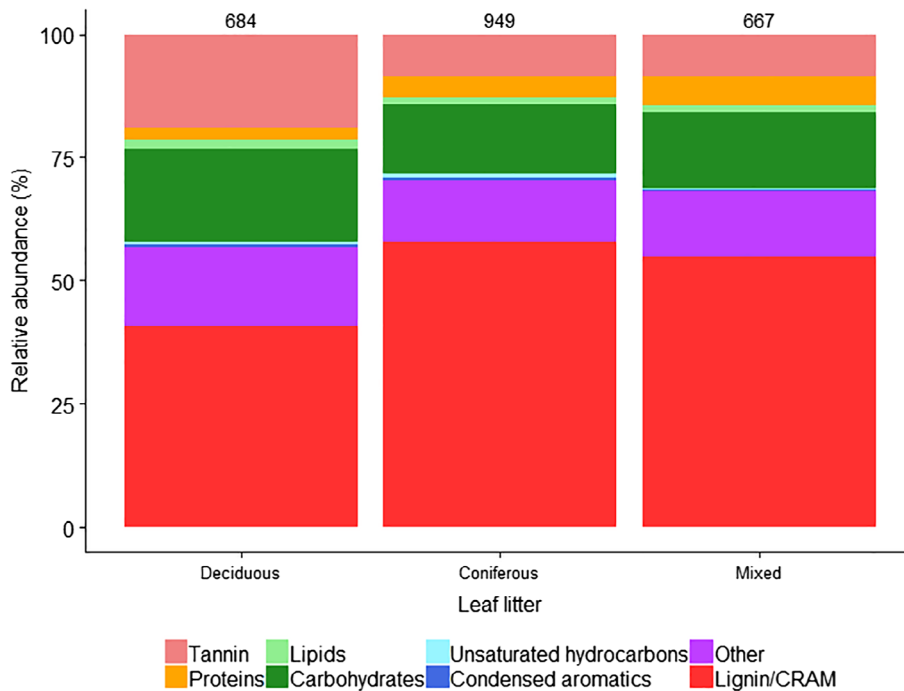


FIGURE 6 Van Krevelen plot (top) and carbon versus mass plot (bottom) for the leaf litter samples. Each dot represents an annotated molecular formula. Formulae found for multiple samples are overlapping. H/C, hydrogen/carbon ratio; O/C, oxygen/carbon ratio

that m/z values ranged from approximately 100 Da to 450 Da, with the highest density around 100–300 Da (Figure 4). More than 75% of the molecular formulae assigned to SRFA consisted of just C, H and O atoms (Figure S1 in File S1) and 82 molecular formulae contained

one S atom (Table S6 in File S1). No molecular formulae containing P atoms remained after all filter steps that are part of our workflow (Table S2 in File S1). The number of feature groups that was found in all four concentrations was small (50) (Figure S2 in File S1), although additional shared

feature groups were found between adjoining concentration samples.

3.3 | Leaf litter

Differences in mean OM composition between the three leaf litter sources – deciduous, coniferous and mixed – were small (Table 1). The number of assigned formulae ranged from 667 for mixed litter to 684 for deciduous litter and 949 for coniferous litter (Table 1). These numbers are higher than the number of formulae assigned to the SRFA samples (151–476). Total organic carbon (TOC) concentrations of the analysed leaf litter samples were approximately 0.2 mg ml^{-1} (compared to an approximate TOC concentration of 0.21 mg ml^{-1} for the lowest-concentration SRFA sample). The O/C and H/C ratios were similar for the three litter sources (Table 1). The O/C ratio was low compared

with the SRFA standard (Table 1). Median DBE was 4 for deciduous litter and 5 for coniferous and mixed litter (Table 1). Most assigned features were classified as lignin/carboxyl-rich alicyclic molecules (CRAM) compounds (Figure 5). The Van Krevelen diagram showed that 53% of the assigned formulae had an O/C value of 0.1–0.6 and an H/C value of 0.6–1.7, which is the region that includes lignin-derived compounds (Figure 6). Median molecular weight was 254 Da for mixed litter, 261 Da for deciduous litter and 266 Da for coniferous litter (Table 1). The carbon versus mass diagram showed most assigned formulae in the 100–400 Da m/z range (Figure 6). Approximately 75% of the molecular formulae assigned to leaf litter contained only C, H and O atoms (Figure S3 in File S1) and 55 formulae contained one S atom (Table S7 in File S1). No formulae with P remained after filtering. The compound class distribution was similar for the three litter sources (Figure 5), despite

TABLE 2 Mean organic matter composition of the leaf litter injections

Sample	MF ^a (<i>n</i>)	C# ^b	H# ^b	O# ^b	O/C ^c	H/C ^c	DBE ^d	MW ^e /Da
Leaf litter (inj. 1)	616	10 (0.23)	12 (0.3)	5 (0.17)	0.55 (0.01)	1.33 (0.02)	4 (0.16)	218 (5)
Leaf litter (inj. 2)	636	10 (0.23)	12 (0.3)	5 (0.16)	0.56 (0.01)	1.38 (0.02)	4 (0.16)	217 (5)
Leaf litter (inj. 3)	633	10 (0.23)	12 (0.3)	5 (0.16)	0.56 (0.01)	1.33 (0.02)	4 (0.16)	210 (5)
Leaf litter (inj. 4)	635	10 (0.22)	12 (0.3)	5 (0.16)	0.55 (0.01)	1.38 (0.02)	4 (0.16)	213 (5)
Leaf litter (inj. 5)	639	10 (0.22)	12 (0.3)	5 (0.16)	0.56 (0.01)	1.38 (0.02)	4 (0.16)	212 (5)

^aNumber (*n*) of annotated molecular formulae.

^bMean numbers of carbon, hydrogen and oxygen numbers. Standard errors of the mean are shown in brackets.

^cAtomic ratios of oxygen and hydrogen to carbon.

^dDouble bond equivalence (Koch & Dittmar, 2006).

^eMolecular weight in Da (median).

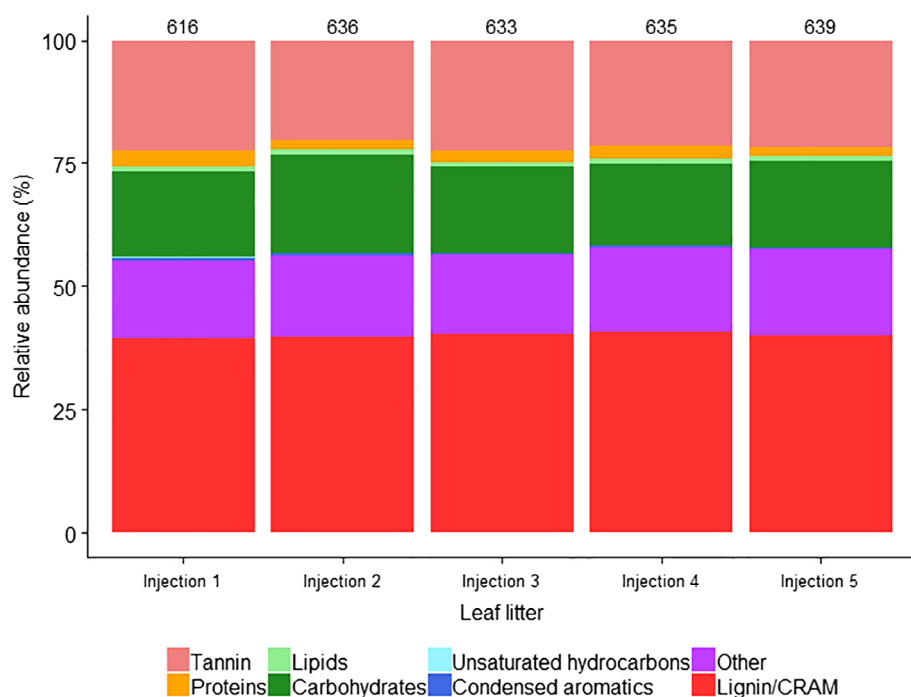


FIGURE 7 Compound class distribution for the leaf litter repeatability test. For the molecular compound class definitions see Table S3 in File S1. CRAM, carboxyl-rich alicyclic molecules

a large part of the annotated feature groups (131–441) being unique for the respective litter source and just 151 feature groups being found in all three litter sources (Figure S4 in File S1). An additional 333 feature groups were found in both mixed and coniferous litter and 52 in both mixed and deciduous litter.

Molecular characteristics were similar for five consecutively measured injections of the same deciduous leaf litter replicate sample (Table 2). The number of assigned molecular formulae ranged from 616 to 639 for the five injections (Table 2). The O/C ratio was 0.55–0.56, the H/C ratio was 1.33–1.38 and DBE was 4 (Table 2). More than 75% of the assigned formulae consisted of only C, H and O atoms (Table S5 in File S1). For each injection some of the assigned formulae were unique (Figure S6 in File S1), but these were predominantly assigned to low intensity peaks. In total, 385 feature groups were found in all injections and 60 additional feature groups were only absent in one of the five injections. Median molecular weight ranged from 210 Da to 218 Da (Table 2), in line with the m/z range as shown in the carbon versus mass diagram (Figure S7 in File S1). The molecular class distribution was similar for the five injections and was dominated by lignin/CRAM, tannin and carbohydrates (Figure 7).

4 | DISCUSSION

4.1 | Method validation

The method validation test showed that 23 out of 26 naturally occurring leaf litter DOM molecules were correctly identified as features from a complex natural DOM matrix signal when using LC-QTOF-MS in combination with the patRoom R package. Moreover, it was found that the six initially wrongly assigned formulae contained four N atoms and one P atom. However, such formulae with the current settings are not considered as they are difficult to resolve even with the best FT-ICR-MS instruments (Hawkes et al., 2016). Filtering out these candidate formulae before selecting the best scoring candidate formula thus resulted in a more accurate formula assignment. We recommend further method validation tests using a more diverse mix of known natural DOM compounds, as the current mix largely consisted of lignin and lipid compounds (Table S5 in File S1).

4.2 | Concentration effects

The greater number of formulae assigned to higher Suwannee River fulvic acid (SRFA) concentrations (Table 1) is most likely due to some features being above the intensity threshold only for the higher concentrations (2.5–5.0 mg ml⁻¹). A previous LC-Orbitrap-MS study (Petras

et al., 2017) also reported a higher number of features with assigned formulae (mass error < 5 ppm) with increased concentration of the injected sample, confirming that the sensitivity of the mass spectrometer was concentration dependent. The features present only in the higher-concentration samples were mainly classified as lignin/CRAM and tannin compounds (Figure 2). Thus, sample concentration affects the observed OM composition, with more lignin/CRAM and tannin compounds found at higher concentrations. Nevertheless, the mean OM composition was similar for the different concentrations (Table 1), except for slightly higher DBE values and molecular weights for the higher concentration samples. The observation that some features are only above the intensity threshold for the higher SRFA concentrations may also explain the low number of feature groups present in all four concentration samples (Figure S2 in File S1).

4.3 | Repeatability

Compound class distributions and mean OM composition were similar for five leaf litter DOM injections of the same replicate sample (Figure 6), despite 37.5–39.7% of the annotated feature groups being absent in at least one of the five litter injections. Further analysis showed that the unique feature groups predominantly had peak intensities just above the intensity threshold and were filtered out for some of the injections if the intensity was just below the threshold. However, we did not consider lowering the intensity threshold in order to retain a good quality/quantity ratio with a high number of features and a low number of features assigned to split peaks or noise (examples in Figure S8 in File S1).

4.4 | Data processing optimization

In the process of optimizing feature extraction it was observed that with stricter settings compounds with higher mass values and with more N, S or P atoms were not included any more. Changing a single parameter used for feature extraction can have a large effect on the total number and quality of the extracted features. For instance, we used a maximum chromatographic peak length window of 240 s. This parameter, called “max_trace_length”, is one of many parameters of the OpenMS “FeatureFinderMetabo” feature-finding algorithm (Kenar et al., 2014) and determines the maximum length of a mass trace. When max_trace_length was turned off the number of extracted features was much larger compared to our settings. However, visual inspection showed that the majority of the features generated with max_trace_length turned off belonged to noise instead of valid chromatographic peaks. Interestingly, differences were small when comparing the OM composition and compound

class distribution for both settings; for example, average molecular weights and elemental ratios remained the same or changed only slightly. This example highlights the importance of data processing optimization when using highly automatized software.

4.5 | Method evaluation and perspectives

The number of assigned formulae for SRFA (151–476) and the 94–474 Da mass range found in this study are comparable to the 220 published molecular formulae and 150–344 Da mass range measured with SEC-QTOF-MS by These et al. (2004). Differences are most likely due to sample heterogeneity and differences in instrumentation and applied methodology.

In this study most intense masses detected using QTOF-MS were in the 100–300 Da mass range, similar to the 100–200 Da mass range in which most intense masses were detected with direct infusion Orbitrap-MS by Remucal et al. (2012). The instrumental settings were tuned for a mass range of 80–1,200 Da. However, also within this optimized m/z range a bias towards lower masses was found. A low bias has also been reported for Orbitrap-MS, with most masses above 500 m/z not being detected (Hawkes et al., 2016). Remucal et al. (2012) provide two explanations for the bias towards the lower mass range compared with FT-ICR-MS. First, the higher intensity of lower mass ions could be due to less effective transmission of higher mass ions through the ESI source. Second, higher mass molecules might fragment into lower mass molecules in the ESI source, even though ESI is considered a gentle ionization source. Fragmentation is caused both by disaggregation of noncovalent or weakly covalently bound molecules and in-source collision-induced dissociation (CID) processes due to energetic collisions between ions and residual gas molecules (Remucal et al., 2012). A third explanation, at least for TOF-MS, might be saturation of the detector. High abundance of closely related travel times will cause the detector to not sufficiently recover between the incidents. Although it is important to be aware of the low mass bias of QTOF and Orbitrap instruments, the majority of the high-molecular-weight molecules may fragment into low-molecular-weight subunits (These et al., 2004) and therefore the mean DOM composition is likely to be largely correct.

The O/C ratio of 0.67–0.82 found for SRFA was slightly higher than the 0.62 based on elemental analysis reported by the IHSS (International Humic Substances Society, 2017). Our H/C ratios of 1.00–1.33, lowest for the highest concentration, were slightly larger than the 0.99 reported by the IHSS. The average H/C ratio reported by These et al. (2004) was 1.2, also higher than the value reported by the IHSS. The slightly higher H/C ratios found in our study and

previous studies may be due to a bias of QTOF-MS (and Orbitrap-MS) towards lower molecular weight (i.e., smaller) compounds. When, as expected, more aromatic compounds are found among high-molecular-weight molecules, a bias towards lower masses results in relatively high H/C ratios. The median DBE values found for SRFA were 3–5, which is in line with the earlier SEC-QTOF-MS study by Reemtsma and These (2005). They reported that only 50% of the molecular formulae provided the 4 or more carbon–carbon double bond equivalents (DBE) required for an aromatic system. Our and previous results indicate that most of the assigned formulae are relatively small and confirm the low mass bias of QTOF-MS instruments. Lastly, the large percentage of lignin/CRAM and tannin compounds observed for SRFA is in line with earlier measurements using FT-ICR-MS (Cory et al., 2010) and Orbitrap-MS (Patriarca et al., 2018; Remucal et al., 2012).

The allowed mass error is an important parameter in the determination of molecular formulae. In this study a mass error of 5 ppm was allowed, which was also used in previous studies (Petras et al., 2017; Rathgeb et al., 2017). Most studies use a scoring system similar to that used in this study to select the best matching molecular formula out of numerous candidate formulae (Petras et al., 2017; Rathgeb et al., 2017). These scores are often based on lowest mass error, isotopic patterns and MS/MS fragments. The method validation tests showed that for our study the scoring-based selection of best matching molecular formulae was correct for the majority of the assigned features (Table S5 in File S1). Selection of the best matching formula is required for visualization (e.g., in a Van Krevelen diagram). The selection of elements to be included in the calculation of candidate formulae is also very important, but is limited by the resolution of the used instrument. As stated by Hawkes et al. (2016), certain arrangements of heteroatoms (Table S2 in File S1), including formulae with P, should not be considered because they cannot be resolved by current high-resolution MS instruments. Lower intensity molecules as well as (most) N, S and P-containing formulae are only resolved at higher resolutions (Hawkes et al., 2016). In most Orbitrap-MS and QTOF-MS studies P atoms are therefore not taken into account (Patriarca et al., 2018; Petras et al., 2017), and when they are taken into account they are often filtered out or not considered (this study and Hawkes et al., 2016). Future improvements in resolution or analysis methods may resolve the problems around the inclusion of certain atoms such as P, Na and Cl.

Rathgeb et al. (2017) analysed riverine DOM and stated that the relatively low number of assigned formulae using LC-QTOF-MS compared to direct infusion FT-ICR-MS is due to a lower resolution and to selectivity of the stationary phase. The advantage of LC-QTOF-MS, however, is the

combination of transient signals with accurate mass spectrometry that facilitates the alignment of ions belonging to one molecular feature (Rathgeb et al., 2017). This makes LC-QTOF-MS a high-resolution alternative to FT-ICR-MS and Orbitrap-MS that can be used for a highly detailed DOM characterization. Benefits of LC-QTOF-MS are the fast scan rate and the online coupling with LC (Rathgeb et al., 2017). Our method combines these advantages of LC-QTOF-MS with a largely automated approach to data analysis using open source and instrument-independent analysis software (patRoom). Our method could be further improved by using molecular fragment data to gain better insight into structural properties.

5 | CONCLUSIONS

We developed and evaluated a novel approach to the characterization of DOM extracted from leaf litter, which includes non-target screening using liquid chromatography (LC) coupled to high-resolution QTOF-MS. Our method is unique because data analysis is largely automated through the use of an in-house-developed R package patRoom. This R package is openly accessible and not restricted to QTOF-MS vendor software. A method validation test showed that out of 26 spiked naturally occurring DOM molecules in a leaf litter DOM sample 23 were retrieved as features. After implementation of additional filtering to exclude certain combinations of heteroatoms, including molecular formulae with P atoms, 22 molecular formulae were correctly assigned. Although a large part of the annotated molecular formulae was unique for the respective litter sources, the molecular compound class distribution was similar for deciduous, coniferous and mixed leaf litter DOM. Most intense masses were detected in the 100–300 Da mass range, showing a bias of QTOF-MS towards lower masses, as earlier reported for Orbitrap-MS. This study shows that LC-QTOF-MS is a high-resolution alternative to Orbitrap-MS and FT-ICR-MS and can be used for non-target screening of leaf litter-derived DOM.

ACKNOWLEDGEMENTS

This study was part of the SoSEAL project funded by the Netherlands Organization for Scientific Research (NWO) and commercial project partners. We thank all commercial partners for providing the scientific freedom to develop this method within the larger SoSEAL project. We thank Jan van Mourik, Jiani Zhou and Susanne Laumann for helping with sampling and general support. We thank Rutger, Chiara, Joke, Samira and Jorien for their kind help in the IBED laboratory of the University of Amsterdam (UvA). We thank Robert Sluijter and Lisa Kooij (both UvA students) for their

experimental work during the environmental measuring technique course. Lastly, we thank an anonymous reviewer and the associate editor for their critical review and suggestions for improvements.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Olaf Brock  <https://orcid.org/0000-0003-4727-8459>

REFERENCES

- Abdulla, H. A. N., Sleighter, R. L., & Hatcher, P. G. (2013). Two dimensional correlation analysis of Fourier transform ion cyclotron resonance mass spectra of dissolved organic matter: A new graphical analysis of trends. *Analytical Chemistry*, *85*, 3895–3902.
- Cory, R. M., McNeill, K., Cotner, J. B., Amado, A. M., Purcell, J. M., & Marshall, A. G. (2010). Singlet oxygen in the coupled photo- and biochemical oxidation of dissolved organic matter. *Environmental Science & Technology*, *44*, 3683–3689.
- Hawkes, J. A., Dittmar, T., Patriarca, C., Tranvik, L., & Bergquist, J. (2016). Evaluation of the Orbitrap mass spectrometer for the molecular fingerprinting analysis of natural dissolved organic matter. *Analytical Chemistry*, *88*, 7698–7704.
- Helmus, R. 2018. R-package: patRoom - hyPhenated mAss specTROmetry nOn-target aNalysis. Retrieved from <https://github.com/rickhelmus/patRoom>.
- International Humic Substances Society. 2017. Elemental compositions and stable isotopic ratios of IHSS samples. Retrieved from <http://humic-substances.org/elemental-compositions-and-stable-isotopic-ratios-of-ihss-samples/>.
- Kaiser, K., & Kalbitz, K. (2012). Cycling downwards – Dissolved organic matter in soils. *Soil Biology and Biochemistry*, *52*, 29–32.
- Kenar, E., Franken, H., Forcisi, S., Wörmann, K., Häring, H.-U., Lehmann, R., ... Kohlbacher, O. (2014). Automated label-free quantification of metabolites from liquid chromatography–mass spectrometry data. *Molecular & Cellular Proteomics*, *13*, 348–359.
- Kim, S., Kramer, R. W., & Hatcher, P. G. (2003). Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the Van Krevelen diagram. *Analytical Chemistry*, *75*, 5336–5344.
- Koch, B. P., & Dittmar, T. (2006). From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry*, *20*, 926–932.
- Koch, B. P., Dittmar, T., Witt, M., & Kattner, G. (2007). Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Analytical Chemistry*, *79*, 1758–1763.
- Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T., & Kattner, G. (2005). Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochimica et Cosmochimica Acta*, *69*, 3299–3308.
- Kujawinski, E. B., & Behn, M. D. (2006). Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance

- mass spectra of natural organic matter. *Analytical Chemistry*, 78, 4363–4373.
- Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, 528, 60–68.
- Libiseller, G., Dvornak, M., Kleb, U., Gander, E., Eisenberg, T., Madeo, F., ... Magnes, C. (2015). IPO: A tool for automated optimization of XCMS parameters. *BMC Bioinformatics*, 16, 1–10.
- Meringer, M., Reinker, S., Zhang, J., & Muller, A. (2011). MS/MS data improves automated determination of molecular formulas by mass spectrometry. *Communications in Mathematical and in Computer Chemistry*, 65, 259–290.
- Nebbio, A., & Piccolo, A. (2013). Molecular characterization of dissolved organic matter (DOM): A critical review. *Analytical and Bioanalytical Chemistry*, 405, 109–124.
- Patriarca, C., Bergquist, J., Sjöberg, P. J. R., Tranvik, L., & Hawkes, J. A. (2018). Online HPLC-ESI-HRMS method for the analysis and comparison of different dissolved organic matter samples. *Environmental Science & Technology*, 52, 2091–2099.
- Petras, D., Koester, I., Da Silva, R., Stephens, B. M., Haas, A. F., Nelson, C. E., ... Dorrestein, P. C. (2017). High-resolution liquid chromatography tandem mass spectrometry enables large scale molecular characterization of dissolved organic matter. *Frontiers in Marine Science*, 4, 1–14.
- R Core Team. (2017). R: A language and environment for statistical computing. Retrieved from <https://www.r-project.org/>
- Rathgeb, A., Causon, T., Krachler, R., & Hann, S. (2017). From the peat bog to the estuarine mixing zone: Common features and variances in riverine dissolved organic matter determined by non-targeted analysis. *Marine Chemistry*, 194, 158–167.
- Reemtsma, T. (2009). Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry: Status and needs. *Journal of Chromatography A*, 1216, 3687–3701.
- Reemtsma, T. (2010). The carbon versus mass diagram to visualize and exploit FTICR-MS data of natural organic matter. *Journal of Mass Spectrometry*, 45, 382–390.
- Reemtsma, T., & These, A. (2005). Comparative investigation of low-molecular-weight fulvic acids of different origin by SEC-Q-TOF-MS: New insights into structure and formation. *Environmental Science & Technology*, 39, 3507–3512.
- Remucal, C. K., Cory, R. M., Sander, M., & McNeill, K. (2012). Low molecular weight components in an aquatic humic substance as characterized by membrane dialysis and Orbitrap mass spectrometry. *Environmental Science & Technology*, 46, 9350–9359.
- Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., ... Kohlbacher, O. (2016). OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nature Methods*, 13, 741–748.
- Scheel, T., Haumaier, L., Ellerbrock, R. H., Rühlmann, J., & Kalbitz, K. (2008). Properties of organic matter precipitated from acidic forest soil solutions. *Organic Geochemistry*, 39, 1439–1453.
- Schmidt, M. W. I., Tom, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., ... Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, 478, 49–56.
- Schmitt-Kopplin, P., Gabelica, Z., Gougeon, R. D., Fekete, A., Kanawati, B., Harir, M., ... Hertkorn, N. (2010). High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 2763–2768.
- Sleighter, R. L., & Hatcher, P. G. (2007). The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *Journal of Mass Spectrometry*, 42, 559–574.
- These, A., Winkler, M., Thomas, C., & Reemtsma, T. (2004). Determination of molecular formulas and structural regularities of low molecular weight fulvic acids by size-exclusion chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 18, 1777–1786.
- Tziotis, D., Hertkorn, N., & Schmitt-Kopplin, P. (2011). Letter: Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform options for the assignment of elemental compositions and the classification of organic molecular intensity. *European Journal of Mass Spectrometry*, 17, 415–421.
- Verkh, Y., Rozman, M., & Petrovic, M. (2018). A non-targeted high-resolution mass spectrometry data analysis of dissolved organic matter in wastewater treatment. *Chemosphere*, 200, 397–404.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Brock O, Helmus R, Kalbitz K, Jansen B. Non-target screening of leaf litter-derived dissolved organic matter using liquid chromatography coupled to high-resolution mass spectrometry (LC-QTOF-MS). *Eur J Soil Sci*. 2020; 71:420–432. <https://doi.org/10.1111/ejss.12894>