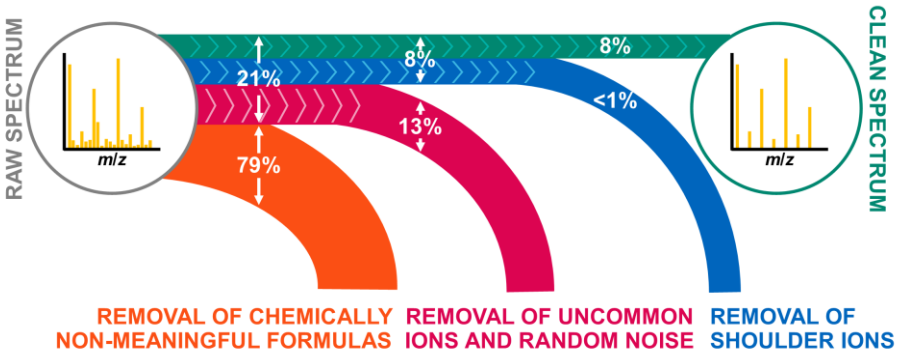
A new processing scheme for ultra-high resolution direct infusion mass spectrometry data

Arthur T. Zielinski[1], Ivan Kourtchev[1], Claudio Bortolini[2], Stephen J. Fuller[1], Chiara Giorio[1], Olalekan A. M. Popoola[1], Sara Bogialli[2], Andrea Tapparo[2], Roderic L. Jones[1], and Markus Kalberer[1*]

[1] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
[2] Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, 35131 Padova, Italy

Abstract

High resolution, high accuracy mass spectrometry is widely used to characterise environmental or biological samples with highly complex composition enabling the identification of chemical composition of often unknown compounds. Despite instrumental advancements, the accurate molecular assignment of compounds acquired in high resolution mass spectra remains time consuming and requires automated algorithms, especially for samples covering a wide mass range and large numbers of compounds. A new processing scheme is introduced implementing filtering methods based on element assignment, instrumental error, and blank subtraction. Optional post-processing incorporates common ion selection across replicate measurements and shoulder ion removal. The scheme allows both positive and negative direct infusion electrospray ionisation (ESI) and atmospheric pressure photoionisation (APPI) acquisition with the same programs. An example application to atmospheric organic aerosol samples using an Orbitrap mass spectrometer is reported for both ionisation techniques resulting in final spectra with 0.8% and 8.4% of the peaks retained from the raw spectra for APPI positive and ESI negative acquisition, respectively.



Highlights

- Ultra-high resolution mass spectrometry processing scheme from acquisition to analysis.
- Method implementable with APPI and ESI ionisation in both polarities.
- Example application for environmental samples showing >90% peak filtering.

Keywords

UHRMS, ESI, APPI, Environmental samples, direct infusion, Orbitrap

* Corresponding author: Markus Kalberer, markus.kalberer@atm.ch.cam.ac.uk

1 Introduction

Obtaining correct elemental composition of unknown molecules by mass spectrometry is a challenge despite advances in instrumentation and data processing algorithms [1]. Ultra-high-resolution mass spectrometry (UHRMS), coupled with soft ionisation techniques, most commonly electrospray ionisation (ESI), can provide a detailed molecular composition for a large, complex sample [2,3] being able to identify many distinct peaks at a given nominal mass. Manual data processing and formula assignment is extremely time consuming [1,4] so automatic algorithms have been developed that generally include noise elimination and blank subtraction steps [5,6]. Noise filtering and blank subtraction is challenging for analysis in direct infusion without prior chromatographic separation as the ion intensities may not be directly related to the concentration of the molecules in the sample [7].

Constraints on allowed chemical elements and number of atoms are used when assigning molecular formulae due to chemical reasons and computational limits. In a molecule containing only carbon and hydrogen the "rule of 13" can be used to limit the number of carbon atoms, in which the nominal mass is divided by 13 and the numerator gives the number of carbon atoms and the remainder gives the number of hydrogens [8,9]. However, natural organic matter is mainly composed of C, H, O and N with minor contributions from S and P, the latter being a quantitatively non-significant component and often not considered [4,10,11]. The number of possible solutions for an elemental formula increases largely if non-oxygen heteroatoms are taken into consideration for calculation. Calculating unique elemental compositions is not always possible [1] when acquiring data with high mass accuracy and resolution, especially as increasing the molecular mass of analytes increases the number of possible molecular formula assignments exponentially [4]. In addition, because mass spectrometry does not directly provide structural information, these molecular formulae may represent any of several structural isomers.

In order to automatically constrain the large number of possible candidates, rules have been developed to select the most likely and chemically meaningful molecular formulae [1]. An important constraint for restricting formulae to those that are likely to exist in nature is including element ratios, especially the H/C ratio which, in most cases does not exceed $H/C > 3$ [1]. Similar restrictions can be put on the O/C (taking acidic polysaccharides as the upper limit for molecular oxygen content $O/C < 1.3$) [4] and other heteroatoms to carbon ratios [11–13]. Additional constraints can be applied based on double bond equivalent (DBE), which indicates the number of rings and double bonds in a molecule and is a measure of the degree of unsaturation in a given compound [10]. Neutral molecules must have a DBE with an integer value [8]. However, the elements N, S, and P may have different valences depending on their chemical environment so constraints based on DBE values need to be used with caution [1]. Formulae are often filtered base on the "nitrogen rule" [1,4,14]. Neutral molecules containing an odd total number of ^{14}N atoms always exhibit an odd nominal mass [4]. The nitrogen rule derives from the fact that chemical elements with even nominal mass have an even valence, while elements with odd mass have an odd valence, with the exception of nitrogen [3]. The majority of data processing methods for Fourier transform ion cyclotron resonance technique (FT-ICR) and Orbitrap™ mass spectrometers check for the presence of isotopes rather than using isotopic ratios for formula assignments [4,10,15]. Other mass spectrometers, *e.g.*

TOF-type, often use isotope patterns for compound identification which tend to provide more reliable assignments compared to FT-ICR and Orbitrap™ mass spectrometers [16].

Once chemically meaningful formulae have been filtered, more than one possible formula per peak may still exist, especially at high *m*/*z*. In order to select the most meaningful formula assignment, two general strategies have been applied: a "best-fit" approach, in which the formula with the closest match between theoretical mass and observed mass is selected, and a "formula extension" approach, in which chemical and structural relationships among compounds are taken into account for formula assignment, *e.g.* by looking for homologous series based on Kendrick mass defects [5,14,17,18]. In the first case, possible incorrect assignments may arise from inaccuracies in the measured masses [19–21]. In the second case, it has previously been observed that, for example, atmospheric oxidation reactions involving S- and N-containing functional groups may lead to a wide variety of products that do not produce homologous series, risking the removal of potentially correct assignments [22].

Most of the methods and currently available algorithms were developed based on ESI and therefore they rely on the assumption that ionisation is accompanied by protonation, deprotonation or adduct formation [20,23,24]. Other ionisation techniques, such as atmospheric pressure chemical ionisation (APCI) and atmospheric pressure photoionisation (APPI) are becoming increasingly common for less polar and apolar organic compounds [20,25]. In APPI, detection of molecular ions (as radical cations or anions) over quasi-molecular ions is common [20,25] so there is a need to develop new algorithms that take into account the formation of molecular ions.

Here we developed a code to filter molecular formula assignments that i) can be applied to different soft-ionisation techniques like ESI, and APPI in both positive and negative ionisation, ii) takes into account formation of molecular ions, quasi-molecular ions and Na adducts, iii) uses a novel method for mass shift and noise estimation and iv) can be used with two different blank subtraction methods. Many steps of the scheme are widely used in mass spectrometry studies, but direct comparisons between methods are difficult as detailed procedures are often not available in the literature. Aspects of the approach described here have been previously applied in studies of environmental [12] and biological samples [26].

2 Pre-processing

The following discussion is based on the use of an Orbitrap™ mass spectrometer (LTQ Orbitrap Velos, Thermo Scientific™, Bremen, Germany) with the proprietary software Xcalibur™ 2.1-3.0 (Thermo Scientific™, Bremen, Germany) henceforth referred to simply as *Xcalibur*. The steps taken, however, apply for the general processing of mass spectra with any spectrometer/software combination. While the choice of mass spectrometer does not affect the presented procedure, the resolution of the spectrometer will influence the accuracy of the final spectra with higher resolutions providing clearer peak separation and higher confidence in the molecular formula assignment [27] which is particularly important for complex ambient measurements [28].

2.1 Data acquisition

Ion transmission and ion collection efficiencies in an Orbitrap™ mass spectrometer strongly depend on the *m*/*z* scan range [29]. Therefore, to avoid the loss of the ions at the high or low end of the selected mass range (*e.g.* *m*/*z* 50-1000), it can be split into several overlapping scan ranges [29]. Each of these scan ranges are independently processed and subsequently recombined. Both sample (*i.e.* the spectrum of interest) and blank (*i.e.* a reference spectrum) spectra are acquired under the same conditions. The type of blank will vary based on application but can include solvent, procedural, and field blanks. The blanks are later used to remove contaminants (Section 3.1.3).

The reproducibility of the peak centroids (the mode of the intensity distribution of an individual peak) and their magnitudes is an important aspect of the direct infusion ESI-UHRMS method, especially for low intensity ions due to competitive ionisation and matrix artefacts. Therefore, besides applying noise threshold corrections (Section 3.1.1), it is also important to consider instrumental replicate measurements. For example, natural organic matter sample replicates are considered reproducible with ESI FT-ICR-MS if a minimum of 67% of threshold-corrected peaks are common among replicates [30]. Additionally, longer acquisition times are desirable as the signal-to-noise ratio improves in proportion to the square root of time [31]. The following discussion therefore assumes multiple replicates (used to filter out uncommon assignments as discussed in Section 3.1.6) and multiple scans to provide a reasonable average (as discussed in Section 2.2).

2.2 Data pre-treatment

Each acquired mass spectra is averaged into one spectrum to reduce the noise level. Molecular assignments are performed using *Xcalibur* software applying constraints on the allowable number of each element, the maximum number of possible formulae to assign, and the maximum mass error.

The restriction on the number of elements varies based on application (see Table 1) but the current procedure permits limits on ^{12}C, ^{13}C, ^{1}H, ^{14}N, ^{16}O, ^{32}S, and ^{34}S for both ionisation polarities. For positive ionisation modes up to one ^{23}Na atom is additionally allowed. The absolute element limits are generally determined by dividing the mass range through the element mass; by using the developed set of formulae that were derived from the National Institute of Standards and Technology (NIST), Wiley mass spectra, and the Dictionary of Natural Products (DNP) entries as discussed in Kind and Fiehn [1]; and considering the presence of oligomeric compounds. For the heavy isotopes of carbon and sulphur, their natural abundances are so low that the relative abundance of molecules with multiple heavy isotopes is low (*e.g.* the natural abundance of molecules with more than two ^{13}C atoms is below detection limits [32]). An *a priori* knowledge of expected elements is important as excluding elements potentially removes correct formula assignments and including additional elements risks calculating unrealistic, but lower mass error, assignments [4].

The first 5 (or more) mathematically possible elemental formulae (depending on mass range and instrument accuracy) with the lowest mass error value within a given mass tolerance (up to ±6 ppm) are exported. A wide mass tolerance is used to cover for the observed non-

systematic mass shift, which seems to be dependent on several factors *e.g.* the sample matrix, the ion intensity of individual ions within this matrix, and the mass range. At later data treatment stages the mass shift is estimated (Section 3.1.2) and subsequently corrected during the main processing to significantly narrow down the mass tolerance. Lock masses can be used during acquisition to reduce (but not remove) the observed mass shift [19,31] to improve formula assignment [7,33].

3 Processing and discussion

After pre-treatment *via Xcalibur* each spectrum requires a number of steps to remove unlikely formula assignments associated with, for example, instrument noise and sample contamination. These checks, amongst others, are included in the data treatment discussed below.

3.1 Data treatment

The overall data treatment procedure can be split into three major stages: pre-treatment and extraction, main processing, and post-treatment. Pre-treatment and extraction consists of the steps using *Xcalibur*, discussed above, as well as the initial extraction of mass shift and noise estimations for both sample and blank spectra. The main processing stage includes all major filter and blank subtraction processes. Finally, post-treatment consists of common ion selection (over several repeated measurements of the samples) and shoulder ion removal. The stages and major steps within each stage are listed in Figure 1. The following sections will describe each step in more detail. Individual rounded rectangles denote separate processing scripts which are predominantly written in Mathematica 10.4 (Wolfram Research Inc., UK), henceforth referred to as *Mathematica*, except for the *Xcalibur* processing step which is manually processed.

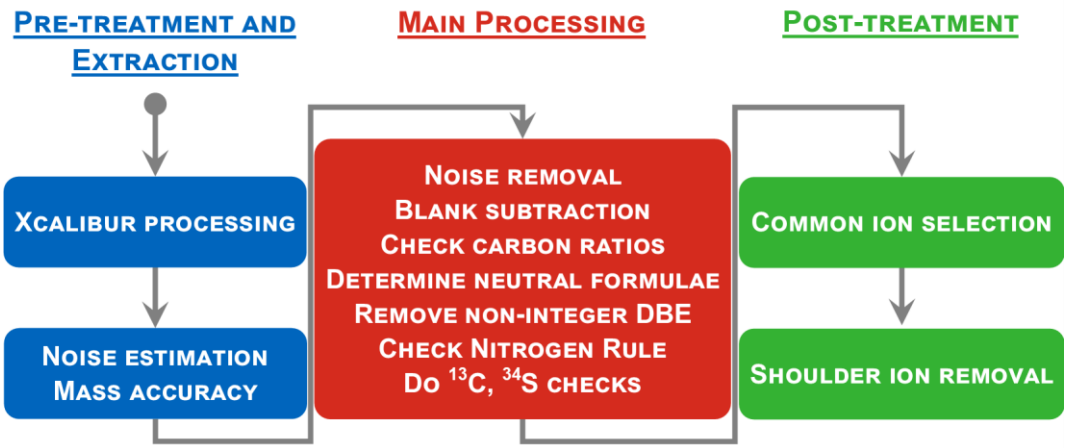


Figure 1 Schematic overview of the three main processing stages of data processing: pre-treatment and extraction (left), main processing (centre), and post-treatment (right). Each rounded rectangle denotes a separate script being used within each stage. The order of steps within the "Main Processing" stage varies slightly based on blank subtraction method (see Section 3.1.3). DBE ≡ double bond equivalent.

3.1.1 Evaluation of noise level

Previous studies have used an extremely wide variety of noise levels evaluated from signal-free regions in the mass spectra [34]. Hawkes *et al.* [35] observed that intensity of noise increases with *m*/*z* in FT-ICR but it is constant in the Orbitrap™ between 150-2000 Da. Signal-to-noise ratios (S/N) ranging between 2 and 20 [2,7,8,10,13,21,24,34,36] have been used which highlights the challenge of reliable noise level estimation in a wide mass range. Other studies simply removed the lowest 10% of peaks (based on intensity) [6] while a S/N of 4 effectively removed peaks with relative intensity below 0.5% in a previous study [24].

The choice of S/N or a cut-off based on intensity seems arbitrary since intensity of peaks is affected by competitive ionisation and cannot be used to infer concentration of compounds when using direct infusion [7,36].

Sleighter *et al.* [30] compared reproducibility in peak detection at different S/N thresholds. They showed that reproducibility increases with increasing S/N cut-off from 3 to 10. However, they point out also that using a strict S/N threshold is not adequate for establishing peak detection reproducibility because well-defined peaks could go undetected just below the defined threshold. They suggest using a lower S/N of 2.5 when looking for common peaks in the other replicates.

In our method, for each processed spectrum (sample and blank), the noise level is estimated based on fitting a normal distribution to a histogram of intensities. The process is visualised in Figure 2 which shows a typical intensity distribution. Histogram bin sizes are selected based on the Freedman-Diaconis rule [37] to ensure the histogram is representative without excessive computer processing. The noise intensity is characterised by a bi-modal normal distribution (Figure 2) which has also been seen by Zhurov *et al.* when describing their alternative approach to the "N sigma" methodology for determining noise levels [38]. The first mode corresponds to the lowest intensity peaks in the MS probably associated with thermal noise [31]. The second mode may correspond to a higher intensity chemical noise [6]. For example, we observe several shoulder ions present nearby to high intensity peaks in the mass spectrum. These artefact ions may have intensities similar to analyte peaks with low concentrations or ionisation efficiency which make up the second mode of the histogram. It is therefore difficult to discriminate between high intensity noise and low intensity analyte peaks.

For this reason, the noise level, which is subsequently used to remove peaks during the main processing stage, is estimated as the mean plus three standard deviations based on the fit of the first mode whenever it is possible to acquire at least three instrumental repetitions. This approach follows the "N sigma" methodology which implements a noise level equal to the mean plus N times the standard deviation (*i.e.* σ). Using $N = 3$ is conservative considering typical N values are 3, 5, 6, or 8 [38]. The second mode is addressed during a later processing stage where only peaks common in all (or a user-defined fraction) of the replicates are kept (see Section 3.1.6) and these are subsequently filtered to remove shoulder ions at masses close to high intensity peaks (see Section 3.1.7).

In contrast, when it is not possible to acquire three instrumental repetitions per sample, *e.g.* in Liquid Extraction Surface Analysis (LESA) when analysing chemically heterogeneous surfaces [26,39], the noise level is estimated as (at least) three times the mean in order to delete all peaks appertaining to the second mode of the intensity distributions (Figure 2) assuming they are high intensity chemical noise.

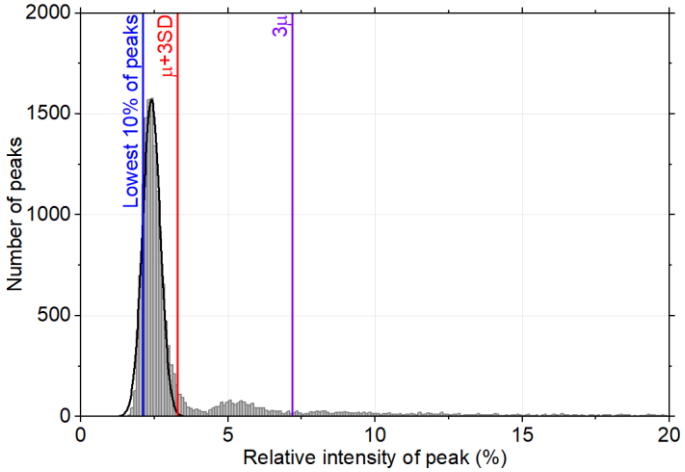


Figure 2 Sample intensity histogram used for noise level estimation using raw data (grey) and fitted curve (black). The noise level (red line) is typically set to the mean plus three standard deviations based on the fit. Under some conditions, when there are not enough replicates for common ion comparisons, the noise level may be set to three times the mean (purple). An additional noise level (blue line) marks the intensity level below which 10% of all peaks reside based on Wong *et al.* (2009) [6]**. The intensities (along the *x*-axis) are given as a percentage of the maximum intensity peak in the spectrum.**

3.1.2 Evaluation of mass shift

The mass shift of each mass spectrum is evaluated based on the measured mass errors of a minimum of ten known reference compounds from *a priori* knowledge (based on the source and polarity used) which are expected in either the sample or solvent. These reference peaks provide an estimate of the overall mass shift of a given spectrum (similar to the method described by Sleighter *et al.* [40]). The reference compounds should be selected to cover the mass range of interest due to the mass dependence of the mass error, as shown using calibration standards in Figure 3, and to be well resolved (or with a higher signal compared to adjacent peaks) in order to avoid inaccuracies in the mass shift evaluation. Mass errors are also dependent on the peak intensity [41] and the matrix but these factors are more difficult to account for and are not considered here systematically.

The errors, as originally calculated by *Xcalibur* in units of ppm, are summarised by their arithmetic mean, standard deviation, maximum, and minimum values to be used in subsequent processing steps. The mass errors are tested for outliers, to avoid skewed data, using Grubbs' Test [42] with a default confidence level of 99%. Any detected outliers are removed from the summary data. While calculating mass shifts, if several peaks are assigned the same chemical formula the highest intensity peak is assumed to have the correct assignment in order to avoid selecting shoulder ions (see Section 3.1.7). Functionally, mass shifts are used as a filter for removing assignments with errors outside an acceptable range (based on the minimum/maximum mass shift of known compounds) effectively restricting the initial wide range set in *Xcalibur*. In order to account for the intra-spectrum mass shift variability mentioned above, a conservative approach was implemented where the maximum and minimum mass shifts allowed for formula assignments can be increased (at the user's discretion) beyond the range observed for the reference compounds (see Figure 3) based on user input with typical values being on the order of 0.5 ppm. The mass error in the example shown in Figure 3 has a range of ca. 2.5 ppm for negative mode (from ca. 0.5 ppm to ca. -2 ppm) and about 4 ppm for positive mode (from ca. 0.5 ppm to ca. -3.5 ppm) ionisation

and illustrates that for most spectra the error is not symmetrically distributed around 0 ppm. At this stage, mass shifts are only used to restrict the range of possible formula assignments without explicitly selecting the correct formula assignment. Mass shifts may also be used during blank subtraction to effectively realign the sample and blank peaks for comparison purposes using the mean and standard deviation as discussed in the following section.

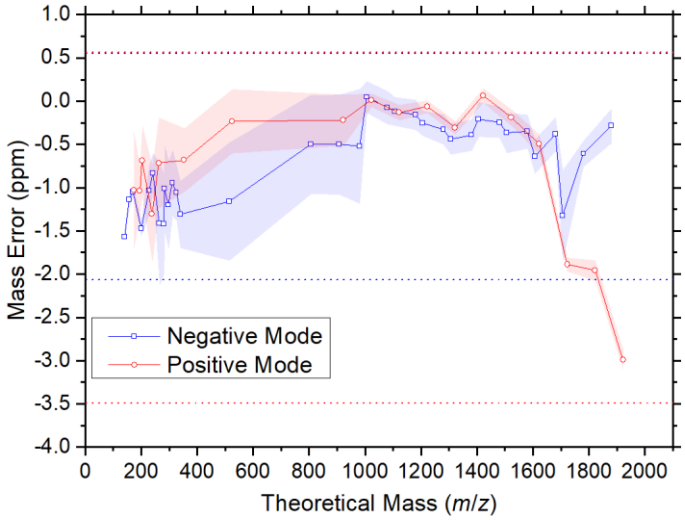


Figure 3 Mass shift of known calibration mixture for ESI positive (red) and negative (blue). Markers denote the mean value from at least three repeats with the shaded area showing the associated standard deviation. Non-uniform, non-linear relationship is shown for both modes with mass errors peaking at extreme *m/z* values. A minimum and maximum approach is used to account for this known relationship as shown with the dashed limit (with an additional offset of 0.5 ppm included). Calibration standards used were the Pierce ESI Negative Ion Calibration Solution (Thermo Scientific) and the Pierce LTQ Velos ESI Positive Ion Calibration Solution (Thermo Scientific). Data taken from three *m/z* scan ranges: 80-600, 150-1000, and 200-2000.

3.1.3 Blank subtraction

Blank subtraction aims to remove any peaks from the sample spectra that are also present in the blank spectra. Two approaches have been studied based on either the final assigned chemical formulae or the corrected masses of each peak. The first method, based on chemical formulae, processes the sample and blank independently and removes matching assigned molecular formulae that have a sample-to-blank ratio below a user-defined limit (*e.g.* 10). The general approach of processing the sample and blank independently before comparison is common in literature [5,43–45], although the removal criteria vary.

The second method checks whether a sample and blank peak are statistically equivalent and removes the sample peak if its intensity is not larger than the sample-to-blank ratio. In this second method the first check, based on a two sample *t*-test, determines whether the sample and blank are within one standard deviation of one another (based on the mass shift standard deviations and correcting for the mean mass shifts to account for inter-spectra variability). If this is true, as demonstrated by blank peaks 1 and 2 in Figure 4, then the second check is performed. The second check determines if the ratio of sample peak intensity to blank peak intensity is below a certain level, *i.e.* the blank is above a certain sample-to-blank limit as illustrated in Figure 4 by a dashed line. If this is true, as in the case of the example blank peak 1, then the sample peak is removed from the dataset. Otherwise, the sample peak is retained

under the assumption that the blank peak is not the same compound (*e.g.* blank peak 3 in Figure 4) or that there is significantly more of the given compound in the sample as compared to the blank (*e.g.* blank peak 2 in Figure 4).

Although both methods produce similar results there are minor differences between the two. The formula-based approach has a strong dependence on the number of assigned formulae during the initial *Xcalibur* processing that may lead to false positives if the same formula is not assigned within the blank spectrum due to the influence of different mass errors on the *Xcalibur* assignment algorithm. That is, for a given number of formula assignments, a small change in measured mass may result in different sets of formulae for the sample and blank. After processing this may allow different final assignments for effectively the same peak resulting in the sample peak being incorrectly kept. On the other hand, the mass-based approach may be limited by the mass error variation. When the standard deviation of the mass shift is high (*i.e.* greater than instrument accuracy), as determined in Section 3.1.2, false negatives due to overzealous blank subtraction may occur. This is due to a wider range of blank peaks being compared to the sample and potentially satisfying the conditions for removal. The result of the difference is typically <5% for the two blank subtraction methods with the formula-based approach having additional false positive assignments with reasonable mass shift variability. Such a comparison, however, is strongly dependent on the number of formulae assigned per peak and the mass shift standard deviation.

The sample-to-blank ratio, when used, is largely arbitrarily selected. Previous studies have used ratios of 10 [2,7,43], up to effectively ∞ [30,45] (*i.e.* everything in the blank is removed from the sample). Rincon *et al.* [44] had a hybrid approach where peaks below a sample-to-blank ratio of 1 would be removed (if within 2 ppm), otherwise the blank intensity was subtracted from the sample assuming the matrix effects were similar for both the sample and blank. For this reason, the sample-to-blank ratio is an adjustable user input during processing.

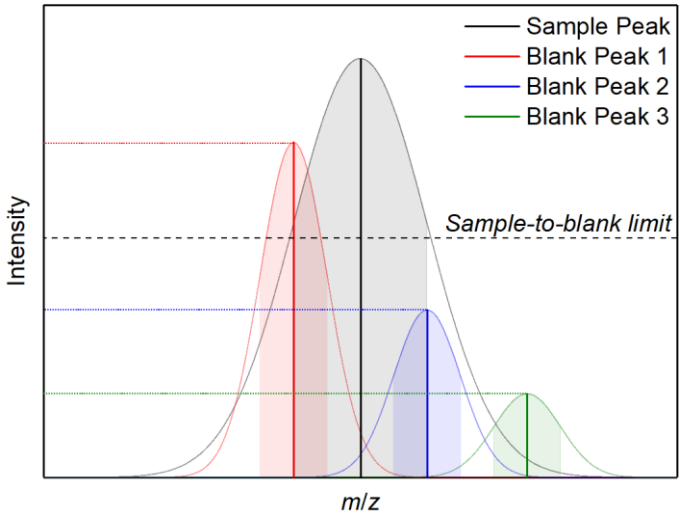


Figure 4 Schematic of the blank subtraction process for a sample peak (black) being compared to three nearby blank peaks (red, green, blue). The first check determines whether the peaks overlap within one standard deviation (*i.e.* shaded regions) which is the case for blank peaks 1 and 2. The second check is whether the sample-to-blank ratio is above a specified sample-to-blank limit which is only true for blank peak 1. Therefore, the sample peak is removed due to the presence of blank peak 1.

3.1.4 Additional exclusion criteria

In addition to noise removal and blank subtraction there are supplementary filters that the main processing code performs.

Carbon ratios

Previous studies (Table 1) have used carbon ratios to eliminate compounds unlikely to naturally exist in the sampled environment. As such, the code allows for control of the exclusion of certain O/C, H/C, N/C, and S/C ratios based on user input. Phosphorous is not currently considered due to its unlikelihood for being a significant component of atmospheric samples (for which this code was initially developed) [10] but is more important for water [46] and soil [11] samples. These limits have seen variability between different references as shown in Table 1 in both value and the choice of ratios used for filtering as expected for varying environments. In the current processing scheme, we refer to the H/C ratio of the neutral molecular formula which is calculated differently depending on ionisation source type (ESI *vs.* APPI) and polarity (positive *vs.* negative). We assume that in ESI the dominant ions are $[M+H]^+$ and $[M+Na]^+$ in positive ionisation and $[M-H]^-$ in negative ionisation. Conversely, in APPI molecular ions $[M]^+$ and $[M]^-$ are also present [25] in addition to quasi-molecular ions and sodium adducts. In order to distinguish between molecular ions ($[M]^+$, $[M]^-$) and quasi-molecular ($[M+H]^+$, $[M-H]^-$) ions we use DBE values. In APPI, when the DBE of the ion is a non-integer we assume it is a quasi-molecular ion while when the DBE is an integer we assume it is a molecular ion.

Table 1 Allowable atom count and carbon ratio ranges from literature to filter out non-naturally existing chemical formula assignments.

		Allowable atom count																		Allowable ratio range									
		^{12}C		^{1}H		^{14}N		^{16}O		^{23}Na		^{32}S		^{34}S		^{31}P		^{13}C		O/C		H/C		N/C	S/C	P/C	(S+P)/C		
Reference	Instrument	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≥	≤	≤	≤	≤	≤		
Koch *et al.* 2005 [8]	FT-ICR	-	100	-	200[1]	-	-	-	50	-	1	-	-	-	-	-	-	-	1	0	1.5	0.3	-	-	-	-	-		
Koch and Dittmar 2006 [47]	FT-ICR	1	100	1	200	0	10	0	50	0	1	-	-	-	-	-	-	0	1	0	1.2	0	2.2	0.5	-	-	-		
Kind and Fiehn 2007 [1][2]	FT-ICR	Allowable atom count varies based on mass range and reference library																		0	1.2	0.2	3.1	1.3	0.8	0.3	-		
Koch *et al.* 2007 [4][3]	FT-ICR	0	∞	0	∞	0	30	0	∞	-	-	0	2	-	-	0	2	-	-	0	1	0.3[4]	-	1	-	-	-		
Wozniak *et al.* 2008 [10][5]	FT-ICR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1.2	0.3	2.25	0.5	0.2	0.1	0.2		
Altieri *et al.* 2009 [48]	FT-ICR	-	∞	-	∞	-	15	-	15	-	-	-	1	-	-	-	1	-	-	0	5[4]	0.3	-	2[4]	-	-	-		
Bateman *et al.* 2009 [45]	LTQ-Orbitrap	-	50[4]	-	100[4]	-	-	-	60[4]	0	1	-	-	-	-	-	-	-	-	0.05	1.3	0.7	2	-	-	-	-		
Schmitt-Kopplin *et al.* 2010 [27]	FT-ICR	-	20	-	30	-	5	-	6	-	-	-	1	-	-	-	-	-	-	0	1	0	2*n*+2	-	-	-	-		
Stubbins *et al.* 2010 [49]	FT-ICR	-	50	2	2*c*+2	-	-	0	*c*+2	-	-	-	-	-	-	-	-	-	-	0	1.2	0.333	2.25	0.5	0.2	0.1	0.2		
Fuller *et al.* 2012 [39]	LTQ-Orbitrap	1	20	-	-	-	-	-	-	0	1	0	1	-	-	-	-	0	1	0	3	0.2	3	1	-	-	-		
Rincón *et al.* 2012 [44]	LTQ-Orbitrap	-	35	-	75	-	7	-	25	-	1	-	7	-	-	-	-	-	1	0	5	0.3	7	6	-	-	-		
Kourtchev *et al.* 2013 [12]	LTQ-Orbitrap	-	100	-	200	-	5	-	50	-	-	-	2	-	1	-	-	-	1	0	1.5	0.3	2.5	0.5	0.2	-	-		
Ohno and Ohno 2013 [11]	FT-ICR	8	50	8	100	0	5	1	30	-	-	0	3	-	-	0	2	-	-	0	1.2	0.3[4]	2.25[4]	0.5[4]	0.2[4]	0.1[4]	0.2		
Kourtchev *et al.* 2014 [7]	LTQ-Orbitrap	-	100	-	200	-	5	-	50	-	-	-	2	-	1	-	-	-	1	0	1.2	0.3	2.5	0.5	0.2	-	-		
Fooshee *et al.* 2015 [50]	LTQ-Orbitrap	1	80	2	140	-	-	0	50	0	1	-	-	-	-	-	-	-	-	0	1.2	0.5	2.2	-	-	-	-		
Lu *et al.* 2015 [13]	FT-ICR	1	50	2	100	0	6	0	30	-	-	0	2	-	-	-	-	-	-	0	1.2	0.35	2.25	0.5	0.2	-	-		
Herzsprung *et al.* 2016 [21]	FT-ICR	-	100	-	-	-	5	-	80	-	-	-	-	-	-	-	-	-	-	0	1	0.3	2*n*+2	-	-	-	-		
Wang *et al.* 2016 [43]	Q-Exactive	1	40	2	80	0	3	0	40	-	-	0	2	-	-	-	-	-	-	0	3	0.3	3	0.5	0.2	-	-		

Lower-bound for ratios is assumed to be zero unless otherwise specified.
Unspecified values denoted by dash (-).
c, *h*, and *n* denote the number of carbon, hydrogen, and nitrogen atoms, respectively.
Inclusion of sodium is for positive ionisation only.
[1] Also includes an additional filtering of $h < 2c + 2$.
[2] Kind and Fiehn 2007 [1] ratio values are covering the 99.7th percentile.
[3] Koch *et al.* 2007 [4] atom count ranges are varied. Quoted values are the most inclusive set.
[4] Quoted value is reported as an exclusive range (*e.g.* < rather than ≤).
[5] Wozniak *et al.* 2008 [10] removed all phosphorus containing compounds after initial filtering.

Double bond equivalent values

Neutral formulae with non-integer (*e.g.* charged molecule [8] and radicals [1]) or negative double bond equivalent (DBE) values are removed from the final peak list. The DBE value is a metric for unsaturation of a given compound based on the number of rings and double bonds [10]. DBEs are calculated using the following:

$$\mathrm{DBE} = 1 - \frac{h}{2} + \frac{n}{2} + \frac{s}{2} + c$$

where *c*, *h*, *n*, and *s* correspond to the number of atoms (*i.e.* C, H, N, and S) in any given chemical assignment $C_cH_hN_nO_oS_s$ [10,12]. Similar to the carbon ratios, the DBE calculation varies based on source and polarity because of the assigned hydrogen count.

Nitrogen rule

Any peaks failing the Nitrogen rule [8], which states that a neutral compound with an odd nominal mass has an odd number of nitrogen atoms [1], are removed.

Carbon and Sulphur isotopes

When assignments contain ^{13}C and/or ^{34}S the filtering process checks for a peak with a matching composition containing only the lighter isotope. If the intensity ratio of heavier-to-lighter isotope was greater than a user-defined factor (*e.g.* 1.2) of the natural isotopic abundance (1.1% and 4.5% for ^{13}C and ^{34}S, respectively [51]) then the isotopic assignment is removed (*i.e.* the assignment with the next smallest mass delta was used). Otherwise, the isotopic assignment is considered the correct assignment and the entire peak is removed as it is chemically equal to the peak containing only the lighter isotope. This process is illustrated in Figure 5.

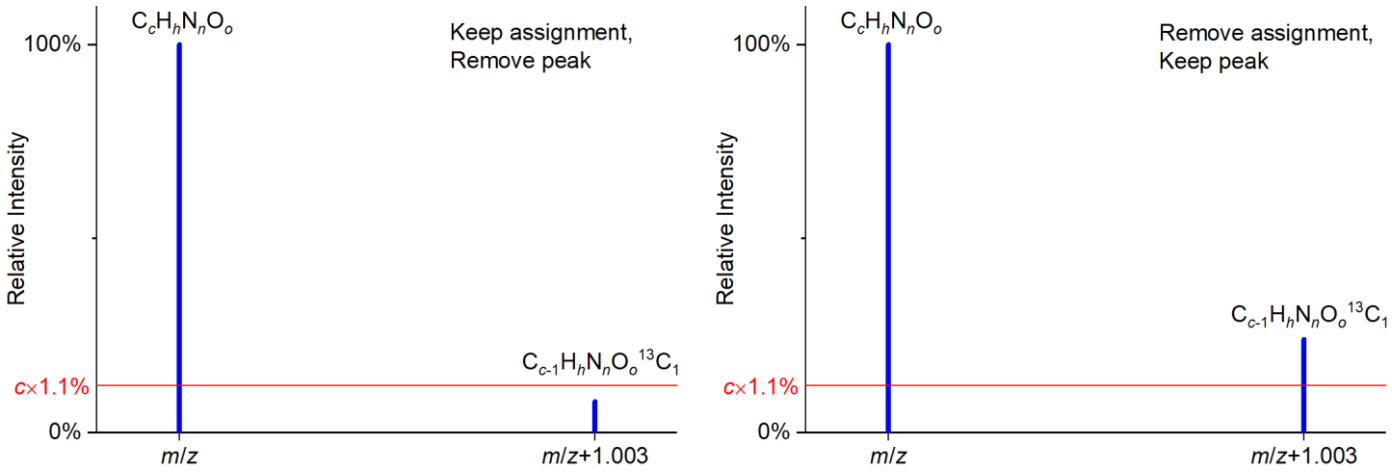


Figure 5 Isotopic removal schematic with $^{13}C/^{12}C$ example showing peaks with ^{12}C isotopes only (left) and with one ^{13}C atom (right) for an arbitrary assignment $C_cH_hN_nO_o$ using a ^{13}C natural abundance of 1.1%. When the relative intensity of the isotope is below (left panel) the $c\times1.1\%$ limit, denoted by the red line, the isotope assignment is considered correct and the peak is removed as it is redundant. Otherwise (right panel), the assignment is deemed incorrect and removed while the peak remains using the assignment with the next smallest mass error.

This approach takes into account the possibility that more than one compound may contribute to the same observed peak within instrumental accuracy and this may be the case when the intensity of the isotopic peak is higher than what expected from its natural abundance. Conversely, the method described by Wozniak *et al.* [10] removed any peak 1.003 *m*/*z* units above another peak under the assumption that the peak at the higher *m*/*z* is always the ^{13}C isotope. Ohno *et al.* [15] removes peaks with intensity lower than 50% of the lighter isotope ion. Heavier-to-lighter isotope ratios tend to be underestimated by both FT-ICR [4] and Orbitrap™ [52] analysers compared to theoretical ratios. Given the likelihood of underestimation, the aforementioned isotopic ratio factor can safely be set to 1 (*i.e.* use the natural abundance) but varying the value can be shown to still influence the number of final peaks – especially if the abundance is lowered (see Figure S1 in Supplementary Material).

3.1.5 Duplicate removals

At this stage of the processing there may still be multiple assignments for a single peak within the derived experimental mass error range (as described in Section 3.1.2). An option within the *Mathematica* script allows for duplicate removal where the assignment with the smallest absolute mass error, after mass shift correction, is kept as the true assignment. The option can be selected at the user's discretion depending on the scenario. In general, if the resolution and accuracy of the instrument used do not allow the identification of a unique formula assignment for a given peak in the mass spectrum, different approaches may be considered. Those include: (i) selecting the assignment with the smallest mass error (option available within the script); (ii) selecting the assignment using a "formula extension" approach based on Kendrick mass defects (not implemented within the script); (iii) keeping all possible assignments (option available within the script if the duplicate removal option is not used); and (iv) removing all peaks for which multiple assignments are still present (possible to do manually after the data processing). Although the first approach is implemented in the processing scheme, care should be taken since the best formula assignment may not always be assignment with the lowest mass error [53]. The "formula extension" approach based on Kendrick mass defects was not implemented because, as mentioned in the introduction, S- and N-containing functional groups do not necessarily form a homologous series in a Kendrick mass defect plot [22].

3.1.6 Common ion selection

After several replicate mass spectra (minimum of 3) of each sample are processed through the main processing stage of Figure 1, they are simultaneously compared for common ions. This common ion selection process filters out any peaks that do not exist in all (or some fraction) of the processed replicates as chosen by the user. The output intensity of the common ions is the average of the replicates. Common ion selection removes the second mode visible in the noise histograms (Figure 2), as discussed above, as those peaks are predominantly noise.

3.1.7 Shoulder peak removal

Shoulder peaks are artefacts of the mass spectrometer's processing produced during the Fourier transform calculation [54]. Figure 6 shows a scenario of shouldering (inset) which is highlighted by a shaded Lorentzian curve fit. The high intensity peak is bordered by several low intensity shoulder peaks. Given the difficulty to identify these artefacts, a conservative approach is used to remove apparent shoulders based on the assumption that shoulders are more

likely as peak intensity increases. If a peak is intense enough (*e.g.* >1,000,000), any neighbouring local ions (*e.g.* within ±0.01 *m/z*) that are less than a specified percentage (*e.g.* 1%) of the local major peak are considered shoulders and removed. More intense peaks are kept as they could still be considered true peaks. The local peak intensity, mass range, and shoulder percentages are adjustable by the user.

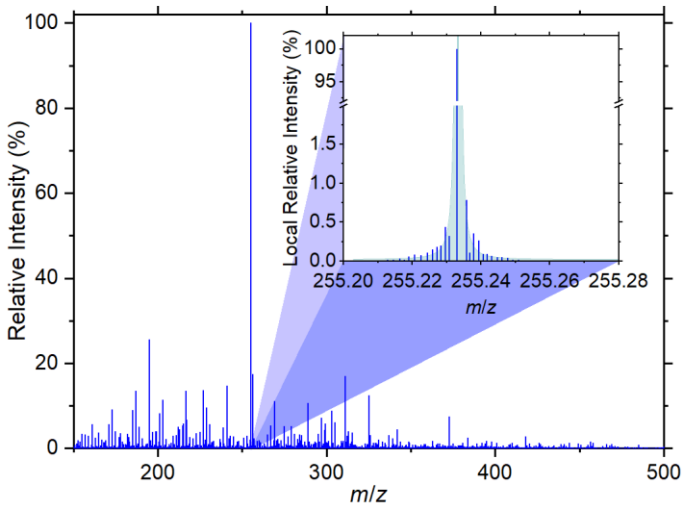


Figure 6 An example of shoulder peaks surrounding a high intensity peak (inset). A shaded Lorentzian fit highlights the shoulder peaks that are all below 1% relative intensity and within ±0.02 *m/z*. The example shown was analysed in the ESI negative ionisation mode.

3.2 Sample Application: UHRMS analysis of the organic fraction of urban $PM_{2.5}$ samples

The sample application of the data processing procedure is based on a $PM_{2.5}$ filter taken on 30th May 2014 over 24 hrs at an urban background site in the city centre of Padua (Italy), located in the polluted Po Valley. More details of the sampling site and procedures for sample collection are reported elsewhere [55]. The filter was extracted in methanol using the procedure already described elsewhere [7] and analysed in both APPI and ESI in both polarities. Here we show the sample processing performed for APPI positive and ESI negative ionisation modes.

3.2.1 Instrumental analysis/Data acquisition

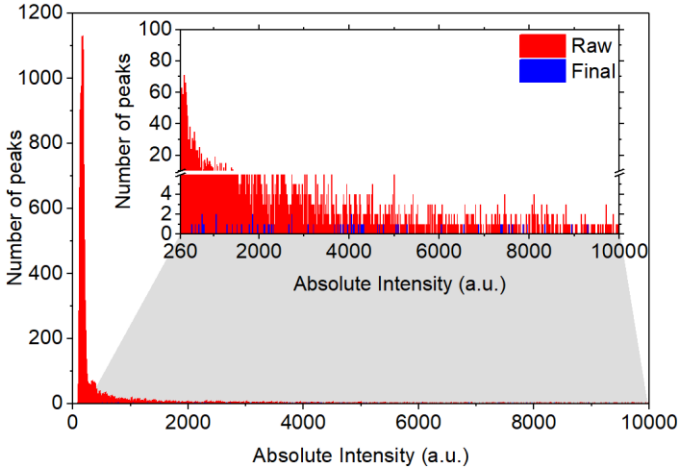
Samples were analysed with a high resolution LTQ Orbitrap Velos mass spectrometer (Thermo Scientific™, Bremen, Germany) equipped with a TriVersa Nanomate® chip-based ESI source (Advion Biosciences, Ithaca NY, USA) and APPI Ion Max source (Thermo Scientific™, Bremen, Germany) with a Syagen Krypton lamp emitting photons at 10.0 eV and 10.6 eV.
The direct infusion negative nanoESI parameters were as follows: ionization voltage -1.6 kV, back pressure 0.8 psi, capillary temperature 275 °C, S-lens RF level 60%, sample volume 8 µL. For analysis in APPI, methanolic extracts doped with 10% toluene were infused at a flow rate of 10 µL/min, with a source temperature of 200 °C, a sheath gas flow of 0 L/min, an auxiliary gas flow of 5 L/min, and a sweep gas flow of 10 L/min. The mass analyser was calibrated before the analysis using Pierce LTQ Velos ESI Positive Ion Calibration Solution (Thermo Scientific) and Pierce ESI Negative Ion Calibration Solution (Thermo Scientific). The mass accuracy of the instrument was checked before the analysis and was below 0.5 ppm. The instrument mass resolution was set at 100,000 (at *m/z* 400). Each sample was analysed in the

m/*z* ranges 100−650 and 150-900, acquiring four repeats for 60 seconds each (~40 scans) in centroid mode.

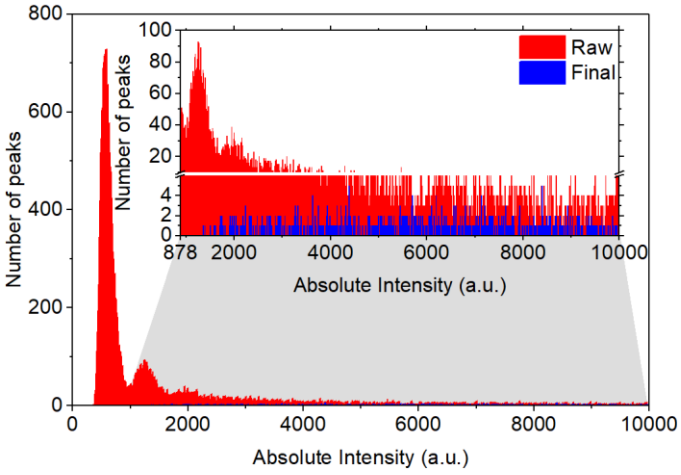
The four sample repeats processed for the *m*/*z* range of 150-900 are discussed here for each ionisation mode. Within *Xcalibur* the chemical assignments were determined for up to 10 formulae per peak allowing a maximum mass error of ±6 ppm. The formula assignment was performed using $1 \leq {}^{12}C \leq 75$, $0 \leq {}^{13}C \leq 1$, $1 \leq {}^{1}H \leq 180$, $0 \leq {}^{16}O \leq 50$, $0 \leq {}^{14}N \leq 30$, $0 \leq {}^{32}S \leq 2$, $0 \leq {}^{34}S \leq 1$. The positive ionisation mode additionally allowed for up to one sodium atom.

3.2.2 Data processing

The data from each ionisation mode was processed using a noise level based on the mean plus three standard deviations definition (as discussed in Section 3.1.1). The resulting noise levels were calculated to be 260 and 878 for the APPI positive and ESI negative modes, respectively. A comparison of the intensity histograms before and after the entire processing are shown in Figure 7 as a means of examining the change in high count, low intensity peaks typically associated with noise. As expected, the secondary mode was removed during the common ion stage of post-processing and the entire histogram was reduced to sub-20 counts for all intensities.



(a)



(b)

Figure 7 Intensity histograms before (red) and after (blue) processing for (a) APPI+ and (b) ESI− samples. Inset zooms to the range above the preset 'noise limit' calculated during pre-processing (260 and 878, respectively). The secondary mode was removed throughout the processing highlighting the effectiveness of common ion filtering.

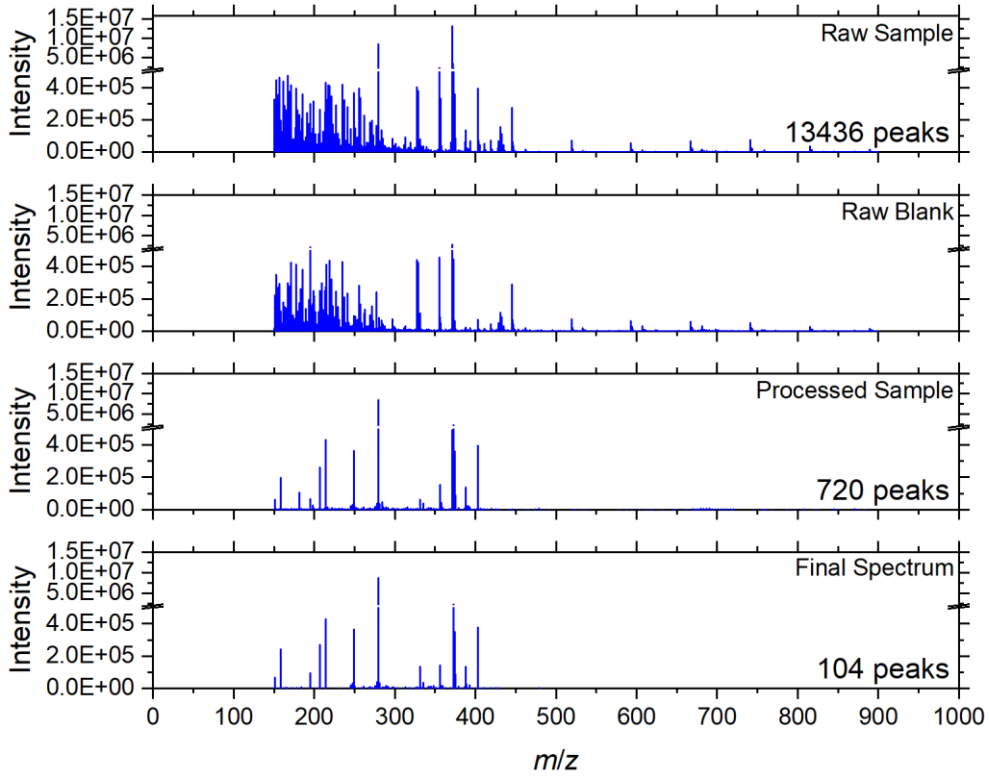
The main processing stage used a sample-to-blank ratio of 5 for blank subtraction with the mass-based approach. Allowable carbon ratios were set to $0.3 \leq H/C \leq 2.5$, $0 \leq O/C \leq 2$, $N/C \leq 1.3$, $S/C \leq 0.8$ and natural abundances were used for the carbon and sulphur isotopic ratios. The results of the main processing stage are shown in Figure 8 for both ionisation modes. A clear reduction in peaks is visible which is largely due to noise removal and blank subtraction, along with the additional filters, bringing the total assignment counts from 87,217 and 238,006 to 720 and 6,491 for the APPI positive and ESI negative modes, respectively. The assignment count at this stage is equivalent to the peak count as duplicates were removed. Any peaks without assignments were removed from the spectra prior to analysis. The total assignment counts at the various stages of the data processing procedure, discussed further below, are summarised in Table 2. These values are specific to the current sample analysis and may not be representative for different analyses especially when adjusting the number of allowable formula assignments.

Table 2 Number of assignments at various stages of processing, following the breakdown in Figure 1, given in absolute (Abs.) and relative (Rel.) terms. The assignment counts prior to common ion selection were based on a single repeat (same as in Figures 7 and 8). Common ion selection was based on four repeats for each ionisation mode. Duplicate assignments for a given peak were removed. These results are for the sample application only and may not be representative for different samples. The final number of assignments remaining correspond to 0.8% and 4.8% of the initial number of peaks for APPI+ and ESI-, respectively.

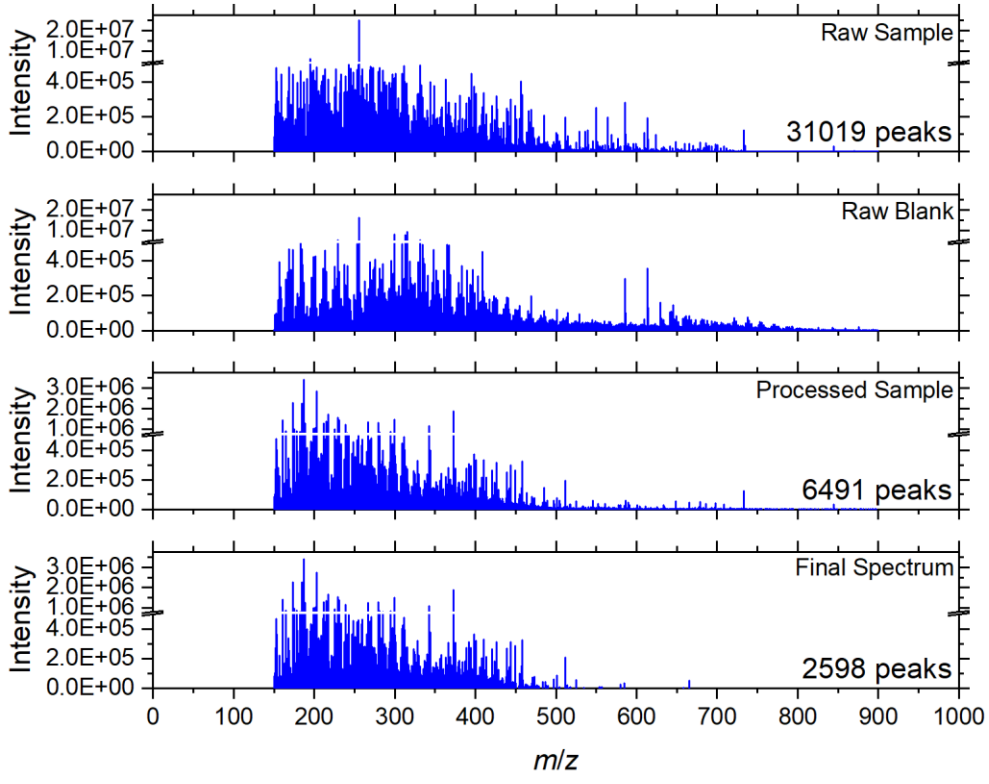
		APPI+		ESI-	
Completed processing		Abs.	Rel.	Abs.	Rel.
Pre-treatment		87,217	100.0%	238,006	100.0%
Main processing	*Noise removal*	33,247	38.1%	123,757	53.3%
	Blank subtraction	17,468	20.0%	100,343	43.2%
	Filter mass error range	9,086	10.4%	78,999	34.0%
	Filter carbon ratios	5,311	6.1%	46,217	19.9%
	DBE and nitrogen rule	2,978	3.4%	23,812	10.3%
	Isotopic filtering	1,006	1.2%	7,681	3.3%
	Duplicate removal [a]	720	0.8%	6,491	2.8%
Common ion selection of four repeats		105	0.1%	2,603	1.1%
Shoulder ion removal		104	0.1%	2,598	1.1%

[a] Removal of duplicate assignments is an optional processing step.

The four repeats of each ionisation mode were processed under the same conditions and subsequently processed to check for common ions. The common ion processing specified that a peak must appear in all four repeats to remain. This stage left 105 and 2,603 peaks for the APPI positive and ESI negative modes, respectively, before being checked for shoulder ions. Peaks were considered to be shoulders if their intensity was less than 1% of a high intensity peak (>1,000,000 a.u.) within 0.01 *m*/*z*. Only one and five peaks were considered shoulders for the APPI positive and ESI negative modes, respectively, largely due to the shoulder peaks being removed in earlier processing steps. The final spectrum for each ionisation mode is presented as the final panel in Figure 8. A more detailed description of shoulder ion removal is included in the Supplementary Material (Figure S2). The processing of the two sample sets concludes with 104 and 2,598 non-duplicate assignments remaining for the APPI positive and ESI negative modes, respectively.



(a)



(b)

Figure 8 Mass spectra of the raw sample (first panel), blank (second panel), and processed sample (third panel) of a single replicate for the (a) APPI+ and (b) ESI− ionisation modes. The fourth panel is the final spectrum across four replicates after common ion retention and shoulder ion removal. Total peak numbers decreased from 13,436 to 104 and 31,019 to 2,598 for APPI+ and ESI−, respectively, with duplicate assignments removed. Note the raw sample peaks had multiple assignments (see Table 2).

4 Conclusions

In order to obtain realistic formula assignment from UHRMS data we introduced a processing scheme based on knowledge of the samples being analysed and the instrument itself. The developed scheme can be implemented for ESI and APPI ionisation techniques in both positive and negative modes.

The procedure implements several common steps including *a priori* element selection and exclusion filtering. Exclusion filtering, typically based on prior knowledge, includes known instrument errors, general chemical principles (*e.g.* the nitrogen rule), and assumptions on realistic element ratios. Pre-treatment considerations are also included to determine the noise level and mass shift across each averaged spectrum.

Two methods of blank subtraction are available based on either processing the sample and blank independently before comparison or performing the subtraction during the main processing stage. While the former approach is more common, the latter approach typically results in fewer false positives. The final spectrum also undergoes common ion selection to exclude chemical noise peaks, when at least three replicates are acquired, and shoulder ion removal for high intensity peaks.

A sample application of the processing scheme was presented, using both ionisation techniques, to highlight the effectiveness of each stage in terms of peak removal. The final peak counts were 0.8% and 8.4% for APPI positive and ESI negative ionisation, respectively, relative to the number of peaks in the initial raw spectra with duplicate peak assignments removed.

Potential future improvement of the overall processing scheme may involve the inclusion of m/z-dependent mass shifts and additional elemental assignments such as phosphorus. The current iteration, however, has already shown to be capable of processing complex atmospheric compositions [12,22]. The approach allows for significant user-input enabling a wide range of potential sample compositions and sampling methods (*e.g.* liquid extractive surface analysis of flower petals [26]).

Acknowledgement

This work was supported by the European Research Council (ERC starting grant 279405) and by the U.K. Natural Environment Research Council (NERC grant NE/H52449X/1). ATZ thanks the Natural Sciences and Engineering Research Council of Canada, the Sir Winston Churchill Society of Edmonton, and the Cambridge Trust for PhD funding. IK was supported by a M. Curie Intra-European fellowship (project no. 254319).

Supplementary material

The codes associated with the processing scheme – named "Direct Infusion Mass Spectrometry Processing (DIMaSP)" – are available for download at https://doi.org/10.17863/CAM.9495, together with a basic manual. The current versions were tested for use with Wolfram Mathematica 11.1.

References

[1] T. Kind, O. Fiehn, Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics. 8 (2007) 105. doi:10.1186/1471-2105-8-105.

[2] I. Kourtchev, I.P. O'Connor, C. Giorio, S.J. Fuller, K. Kristensen, W. Maenhaut, J.C. Wenger, J.R. Sodeau, M. Glasius, M. Kalberer, Effects of anthropogenic emissions on the molecular composition of urban organic aerosols: an ultrahigh resolution mass spectrometry study, Atmos. Environ. 89 (2014) 525–532. doi:10.1016/j.atmosenv.2014.02.051.

[3] S.A. Nizkorodov, J. Laskin, A. Laskin, Molecular chemistry of organic aerosols through the application of high resolution mass spectrometry, Phys. Chem. Chem. Phys. 13 (2011) 3612–3629. doi:10.1039/c0cp02032j.

[4] B.P. Koch, T. Dittmar, M. Witt, G. Kattner, Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter, Anal. Chem. 79 (2007) 1758–1763. doi:10.1021/ac061949s.

[5] G. Danger, F.-R. Orthous-Daunay, P. de Marcellus, P. Modica, V. Vuitton, F. Duvernay, L. Flandinet, L. Le Sergeant d'Hendecourt, R. Thissen, T. Chiavassa, Characterization of laboratory analogs of interstellar/cometary organic residues using very high resolution mass spectrometry, Geochim. Cosmochim. Acta. 118 (2013) 184–201. doi:10.1016/j.gca.2013.05.015.

[6] C.C.L. Wong, D. Cociorva, J.D. Venable, T. Xu, J.R. Yates, Comparison of different signal thresholds on data dependent sampling in Orbitrap and LTQ mass spectrometry for the identification of peptides and proteins in complex mixtures, J. Am. Soc. Mass Spectrom. 20 (2009) 1405–1414. doi:10.1016/j.jasms.2009.04.007.

[7] I. Kourtchev, S.J. Fuller, C. Giorio, R.M. Healy, E. Wilson, I. O'Connor, J.C. Wenger, M. McLeod, J. Aalto, T.M. Ruuskanen, W. Maenhaut, R. Jones, D.S. Venables, J.R. Sodeau, M. Kulmala, M. Kalberer, Molecular composition of biogenic secondary organic aerosols using ultrahigh-resolution mass spectrometry: comparing laboratory and field studies, Atmos. Chem. Phys. 14 (2014) 2155–2167. doi:10.5194/acp-14-2155-2014.

[8] B.P. Koch, M. Witt, R. Engbrodt, T. Dittmar, G. Kattner, Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Geochim. Cosmochim. Acta. 69 (2005) 3299–3308. doi:10.1016/j.gca.2005.02.027.

[9] J.W. Bright, E.C.M. Chen, Mass spectral interpretation using the "rule of '13,'" J. Chem. Educ. 60 (1983) 557. doi:10.1021/ed060p557.

[10] A.S. Wozniak, J.E. Bauer, R.L. Sleighter, R.M. Dickhut, P.G. Hatcher, Technical Note: Molecular characterization of aerosol-derived water soluble organic carbon using ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Atmos. Chem. Phys. 8 (2008) 5099–5111. doi:10.5194/acp-8-5099-2008.

[11] T. Ohno, P.E. Ohno, Influence of heteroatom pre-selection on the molecular formula assignment of soil organic matter components determined by ultrahigh resolution mass spectrometry, Anal. Bioanal. Chem. 405 (2013) 3299–3306. doi:10.1007/s00216-013-6734-3.

[12] I. Kourtchev, S. Fuller, J. Aalto, T.M. Ruuskanen, M.W. McLeod, W. Maenhaut, R. Jones, M. Kulmala, M. Kalberer, Molecular composition of boreal forest aerosol from Hyytiälä, Finland, using ultrahigh resolution mass spectrometry, Environ. Sci. Technol. 47 (2013) 4069–4079. doi:10.1021/es3051636.

[13] Y. Lu, X. Li, R. Mesfioui, J.E. Bauer, R.M. Chambers, E.A. Canuel, P.G. Hatcher, Use of ESI-FTICR-MS to characterize dissolved organic matter in headwater streams draining forest-dominated and pasture-dominated watersheds, PLoS One. 10 (2015) 1–21. doi:10.1371/journal.pone.0145639.

[14] E.B. Kujawinski, M.D. Behn, Automated analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra of natural organic matter, Anal. Chem. 78 (2006) 4363–4373. doi:10.1021/ac0600306.

[15] T. Ohno, R.L. Sleighter, P.G. Hatcher, Comparative study of organic matter chemical characterization using negative and positive mode electrospray ionization ultrahigh-resolution mass spectrometry, Anal. Bioanal. Chem. 408 (2016) 2497–2504. doi:10.1007/s00216-016-9346-x.

[16] G. Glauser, N. Veyrat, B. Rochat, J.-L. Wolfender, T.C.J. Turlings, Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers, J. Chromatogr. A. 1292 (2013) 151–159. doi:10.1016/j.chroma.2012.12.009.

[17] C.A. Hughey, C.L. Hendrickson, R.P. Rodgers, A.G. Marshall, K. Qian, Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra, Anal. Chem. 73 (2001) 4676–4681. doi:10.1021/ac010560w.

[18] A.L. Putman, J.H. Offenberg, R. Fisseha, S. Kundu, T.A. Rahn, L.R. Mazzoleni, Ultrahigh-resolution FT-ICR mass spectrometry characterization of α-pinene ozonolysis SOA, Atmos. Environ. 46 (2012) 164–

172. doi:10.1016/j.atmosenv.2011.10.003.
[19] J. V Olsen, L.M.F. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, M. Mann, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap, Mol. Cell. Proteomics. 4 (2005) 2010–2021. doi:10.1074/mcp.T500030-MCP200.
[20] T. Reemtsma, Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry: status and needs, J. Chromatogr. A. 1216 (2009) 3687–3701. doi:10.1016/j.chroma.2009.02.033.
[21] P. Herzsprung, N. Hertkorn, W. von Tümpling, M. Harir, K. Friese, P. Schmitt-Kopplin, Molecular formula assignment for dissolved organic matter (DOM) using high-field FT-ICR-MS: chemical perspective and validation of sulphur-rich organic components (CHOS) in pit lake samples, Anal. Bioanal. Chem. 408 (2016) 2461–2469. doi:10.1007/s00216-016-9341-2.
[22] I. Kourtchev, R.H.M. Godoi, S. Connors, J.G. Levine, A.T. Archibald, A.F.L. Godoi, S.L. Paralovo, C.G.G. Barbosa, R.A.F. Souza, A.O. Manzi, R. Seco, S. Sjostedt, J.-H. Park, A. Guenther, S. Kim, J. Smith, S.T. Martin, M. Kalberer, Molecular composition of organic aerosols in central Amazonia: an ultra-high-resolution mass spectrometry study, Atmos. Chem. Phys. 16 (2016) 11899–11913. doi:10.5194/acp-16-11899-2016.
[23] E. V Kunenkov, A.S. Kononikhin, I. V Perminova, N. Hertkorn, A. Gaspar, P. Schmitt-Kopplin, I.A. Popov, A. V Garmash, E.N. Nikolaev, Total mass difference statistics algorithm: a new approach to identification of high-mass building blocks in electrospray ionization Fourier transform ion cyclotron mass spectrometry data of natural organic matter, Anal. Chem. 81 (2009) 10106–10115. doi:10.1021/ac901476u.
[24] T.B. Nguyen, A.P. Bateman, D.L. Bones, S.A. Nizkorodov, J. Laskin, A. Laskin, High-resolution mass spectrometry analysis of secondary organic aerosol generated by ozonolysis of isoprene, Atmos. Environ. 44 (2010) 1032–1042. doi:10.1016/j.atmosenv.2009.12.019.
[25] E. de Hoffmann, V. Stroobant, Mass spectrometry: principles and applications, 3rd ed., John Wiley & Sons, Ltd., Chichester, England, 2007.
[26] C. Giorio, E. Moyroud, B.J. Glover, P.C. Skelton, M. Kalberer, Direct surface analysis coupled to high-resolution mass spectrometry reveals heterogeneous composition of the cuticle of Hibiscus trionum petals, Anal. Chem. 87 (2015) 9900–9907. doi:10.1021/acs.analchem.5b02498.
[27] P. Schmitt-Kopplin, A. Gelencsér, E. Dabek-Zlotorzynska, G. Kiss, N. Hertkorn, M. Harir, Y. Hong, I. Gebefügi, Analysis of the unresolved organic fraction in atmospheric aerosols with ultrahigh-resolution mass spectrometry and nuclear magnetic resonance spectroscopy: organosulfates as photochemical smog constituents, Anal. Chem. 82 (2010) 8017–8026. doi:10.1021/ac101444r.
[28] L.R. Mazzoleni, B.M. Ehrmann, X. Shen, A.G. Marshall, J.L. Collett, Water-soluble atmospheric organic matter in fog: Exact masses and chemical formula identification by ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry, Environ. Sci. Technol. 44 (2010) 3690–3697. doi:10.1021/es903409k.
[29] C. Ranninger, L.E. Schmidt, M. Rurik, A. Limonciel, P. Jennings, O. Kohlbacher, C.G. Huber, Improving global feature detectabilities through scan range splitting for untargeted metabolomics by high-performance liquid chromatography-Orbitrap mass spectrometry, Anal. Chim. Acta. 930 (2016) 13–22. doi:10.1016/j.aca.2016.05.017.
[30] R.L. Sleighter, H. Chen, A.S. Wozniak, A.S. Willoughby, P. Caricasole, P.G. Hatcher, Establishing a measure of reproducibility of ultrahigh-resolution mass spectra for complex mixtures of natural organic matter, Anal. Chem. 84 (2012) 9184–9191. doi:10.1021/ac3018026.
[31] A. Makarov, E. Denisov, O. Lange, S. Horning, Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, J. Am. Soc. Mass Spectrom. 17 (2006) 977–982. doi:10.1016/j.jasms.2006.03.006.
[32] S. Kim, R.P. Rodgers, A.G. Marshall, Truly "exact" mass: elemental composition can be determined uniquely from molecular mass measurement at ~0.1 mDa accuracy for molecules up to ~500 Da, Int. J. Mass Spectrom. 251 (2006) 260–265. doi:10.1016/j.ijms.2006.02.001.
[33] A. Tapparo, C. Giorio, L. Soldà, S. Bogialli, D. Marton, M. Marzaro, V. Girolami, UHPLC-DAD method for the determination of neonicotinoid insecticides in single bees and its relevance in honeybee colony loss investigations, Anal. Bioanal. Chem. 405 (2013) 1007–1014. doi:10.1007/s00216-012-6338-3.
[34] M.C.K. Soule, K. Longnecker, S.J. Giovannoni, E.B. Kujawinski, Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter, Org. Geochem. 41 (2010) 725–733. doi:10.1016/j.orggeochem.2010.05.017.
[35] J.A. Hawkes, T. Dittmar, C. Patriarca, L. Tranvik, J. Bergquist, Evaluation of the Orbitrap mass spectrometer for the molecular fingerprinting analysis of natural dissolved organic matter, Anal. Chem. 88 (2016) 7698–7704. doi:10.1021/acs.analchem.6b01624.
[36] R.E. O'Brien, A. Laskin, J. Laskin, S. Liu, R. Weber, L.M. Russell, A.H. Goldstein, Molecular characterization of organic aerosol using nanospray desorption/electrospray ionization mass

spectrometry: CalNex 2010 field study, Atmos. Environ. 68 (2013) 265–272. doi:10.1016/j.atmosenv.2012.11.056.
[37] D. Freedman, P. Diaconis, On the histogram as a density estimator: L2 theory, Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete. 57 (1981) 453–476. doi:10.1007/BF01025868.
[38] K.O. Zhurov, A.N. Kozhinov, L. Fornelli, Y.O. Tsybin, Distinguishing analyte from noise components in mass spectra of complex samples: Where to cut the noise?, Anal. Chem. 86 (2014) 3308–3316. doi:10.1021/ac403278t.
[39] S.J. Fuller, Y. Zhao, S.S. Cliff, A.S. Wexler, M. Kalberer, Direct surface analysis of time-resolved aerosol impactor samples with ultrahigh-resolution mass spectrometry, Anal. Chem. 84 (2012) 9858–9864. doi:10.1021/ac3020615.
[40] R.L. Sleighter, G.A. McKee, Z. Liu, P.G. Hatcher, Naturally present fatty acids as internal calibrants for Fourier transform mass spectra of dissolved organic matter, Limnol. Oceanogr. Methods. 6 (2008) 246–253. doi:10.4319/lom.2008.6.246.
[41] A.N. Kozhinov, K.O. Zhurov, Y.O. Tsybin, Iterative method for mass spectra recalibration via empirical estimation of the mass calibration function for fourier transform mass spectrometry-based petroleomics, Anal. Chem. 85 (2013) 6437–6445. doi:10.1021/ac400972y.
[42] F.E. Grubbs, Procedures for Detecting Outlying Observations in Samples, Technometrics. 11 (1969) 1–21. doi:10.1080/00401706.1969.10490657.
[43] X.K. Wang, S. Rossignol, Y. Ma, L. Yao, M.Y. Wang, J.M. Chen, C. George, L. Wang, Molecular characterization of atmospheric particulate organosulfates in three megacities at the middle and lower reaches of the Yangtze River, Atmos. Chem. Phys. 16 (2016) 2285–2298. doi:10.5194/acp-16-2285-2016.
[44] A.G. Rincón, A.I. Calvo, M. Dietzel, M. Kalberer, Seasonal differences of urban organic aerosol composition - an ultra-high resolution mass spectrometry study, Environ. Chem. 9 (2012) 298–319. doi:10.1071/EN12016.
[45] A.P. Bateman, S.A. Nizkorodov, J. Laskin, A. Laskin, Time-resolved molecular characterization of limonene/ozone aerosol using high-resolution electrospray ionization mass spectrometry, Phys. Chem. Chem. Phys. 11 (2009) 7931–7942. doi:10.1039/b905288g.
[46] R.L. Sleighter, P.G. Hatcher, Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Mar. Chem. 110 (2008) 140–152. doi:10.1016/j.marchem.2008.04.008.
[47] B.P. Koch, T. Dittmar, From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter, Rapid Commun. Mass Spectrom. 20 (2006) 926–932. doi:10.1002/rcm.2386.
[48] K.E. Altieri, B.J. Turpin, S.P. Seitzinger, Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry, Atmos. Chem. Phys. 9 (2009) 2533–2542. doi:10.5194/acp-9-2533-2009.
[49] A. Stubbins, R.G.M. Spencer, H. Chen, P.G. Hatcher, K. Mopper, P.J. Hernes, V.L. Mwamba, A.M. Mangangu, J.N. Wabakanghanzi, J. Six, Illuminated darkness: molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry, Limnol. Oceanogr. 55 (2010) 1467–1477. doi:10.4319/lo.2010.55.4.1467.
[50] D.R. Fooshee, P.K. Aiona, A. Laskin, J. Laskin, S.A. Nizkorodov, P.F. Baldi, Atmospheric oxidation of squalene: molecular study using COBRA modeling and high-resolution mass spectrometry, Environ. Sci. Technol. 49 (2015) 13304–13313. doi:10.1021/acs.est.5b03552.
[51] J.R. de Laeter, J.K. Böhlke, P. De Bièvre, H. Hidaka, H.S. Peiser, K.J.R. Rosman, P.D.P. Taylor, Atomic weights of the elements: review 2000 (IUPAC technical report), Pure Appl. Chem. 75 (2003) 785. doi:10.1351/pac200375060683.
[52] Y. Xu, J.-F. Heilier, G. Madalinski, E. Genin, E. Ezan, J.-C. Tabet, C. Junot, Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-Orbitrap mass spectrometer for further metabolomics database building, Anal. Chem. 82 (2010) 5490–5501. doi:10.1021/ac100271j.
[53] P. Herzsprung, N. Hertkorn, W. von Tümpling, M. Harir, K. Friese, P. Schmitt-Kopplin, Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view, Anal. Bioanal. Chem. 406 (2014) 7977–7987. doi:10.1007/s00216-014-8249-y.
[54] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics. 11 (2010) 395. doi:10.1186/1471-2105-11-395.
[55] C. Giorio, A. Tapparo, M.L. Scapellato, M. Carrieri, P. Apostoli, G.B. Bartolucci, Field comparison of a personal cascade impactor sampler, an optical particle counter and CEN-EU standard methods for PM10, PM2.5 and PM1 measurement in urban environment, J. Aerosol Sci. 65 (2013) 111–120. doi:10.1016/j.jaerosci.2013.07.013.