International Journal of Mass Spectrometry 389 (2015) 26–38

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Methods to extract molecular and bulk chemical information from series of complex mass spectra with limited mass resolution

Harald Stark[a,b,c,*], Reddy L.N. Yatavelli[a,c,1], Samantha L. Thompson[a,c], Joel R. Kimmel[b,d], Michael J. Cubison[d], Puneet S. Chhabra[b,2], Manjula R. Canagaratna[b], John T. Jayne[b], Douglas R. Worsnop[b,e], Jose L. Jimenez[a,c,**]

[a] *Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA*
[b] *Aerodyne Research Inc., Billerica, MA 01821, USA*
[c] *Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA*
[d] *TOFWERK AG, 3600 Thun, Switzerland*
[e] *Department of Physics, University of Helsinki, Helsinki 00014, Finland*

ARTICLE INFO

Article history:
Received 6 February 2015
Received in revised form 3 August 2015
Accepted 10 August 2015
Available online 19 August 2015

Keywords:
Chemical ionization
High-resolution time-of-flight mass spectrometry
Atmospheric measurements
Analysis of mass spectra

ABSTRACT

The resolution of mass spectrometers is often insufficient to conclusively identify all peaks that may be present in recorded spectra. Here, we present new methods to extract consistent molecular and bulk level chemical information by constrained fitting of series of complex organic mass spectra with multiple overlapping peaks. Possible individual peaks in a group of overlapping peaks are identified by both defining a chemical space and by free peak fitting. If simply all possible formulas from the chemical space would be used to fit each peak, the result would not be well constrained. The free peak fitting algorithm provides information about likely peak locations. A new algorithm then reconciles the results of both methods and produces a final peak list for use in subsequent fitting, while using all available experimental constraints. Comparison to ultra-high resolution data suggests that the real peak density is substantially higher than can be resolved with the instrument resolution. Bulk chemical properties such as carbon number (nC) and carbon oxidation state (OS_C) can be calculated from the fit results. For mixtures of compounds dominated by C, H, O and N, bulk properties can be reliably extracted, even though some formula assignments may remain uncertain. This ability to retrieve correct bulk parameters even if not all assigned formulas are correct originates from the relationship between mass defects of individual peaks and the chemical parameters under our CHON composition assumptions. Retrieving consistent bulk parameters across series of many mass spectra is essential for extracting time trends, e.g. for field measurements taking place over several weeks. We illustrate the fitting method using a sample data set from a chemical ionization mass spectrometer with a resolution of approximately 4000 (M/dM), operated using acetate reagent ions. Spectral simulation experiments validate the analysis method by showing good agreement of intensities for many specific ions, as well as for bulk chemical parameters. An alternative method to directly extract bulk chemical information from the raw spectra without the need of any peak assignment or peak fitting is also introduced, which shows good agreement with the peak fitting results. The latter method can be applied very rapidly without the need for complex analysis procedures, e.g. as a quick online diagnostic during data acquisition.

1. Introduction

Mass spectrometry is a commonly used tool in many research fields. When analyzing complex samples without chromatographic separation, such as organic samples from atmospheric aerosols, mass spectra can contain molecular ions from hundreds to thousands of different organic compounds [1–3]. This is especially true for a soft-ionization technique such as chemical ionization [4,5]. If ion peaks are not fully resolved from each other, as is

---

[*] Corresponding author at: Aerodyne Research Inc., Billerica, MA 01821, USA. Tel.: +1 9786639500.
[**] Corresponding author at: Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309, USA. Tel.: +1 3034923557.
E-mail addresses: harald.stark@aerodyne.com (H. Stark), jose.jimenez@colorado.edu (J.L. Jimenez).
[1] Now at California Air Resources Board, El Monte, CA 91731, USA.
[2] Now at PerkinElmer Inc., Hopkinton, MA 01748, USA.

http://dx.doi.org/10.1016/j.ijms.2015.08.011
1387-3806/

inevitably the case at higher mass-to-charge ratios (m/z) [6,7], a spectral fitting method is needed. Such a method will help to extract chemical information from the recorded spectra for either molecular or "bulk" properties averaged across all formulas. The term "molecular" is used here to represent all possible isomers of a specific sum formula, since mass spectrometry can only provide information about individual exact m/z values, but not about functional groups or molecular structure. Bulk properties, such as carbon oxidation state (OS_C) and number of carbon atoms (nC), are useful indicators of atmospheric processing (i.e. oxidation and/or chemical fragmentation) of organic compounds [6,8]. In addition to retrieving this information from individual spectra, it is often as important to get consistent data from multiple mass spectra, as is the case, for example, for field campaigns, where regularly tens of thousands of mass spectra are acquired over weeks and time series of both molecular and bulk parameters are needed.

An example for the large amount of possible chemical formulas present at higher m/z is shown in Fig. 1, which will be discussed in more detail later. The left column shows individual peaks from possible chemical formulas containing C, H, O and N for ions at three different m/z ranges. The number of possible ions increases rapidly with m/z and all possible ions are not resolvable at $m/z > 200$ (bottom row). Assumptions or additional information about the composition of the sample (e.g. elements present) are often useful to limit the number of possible formulas that could be assigned to the ions from the sample. Kind and Fiehn [9] developed a set of rules to determine the most likely formula at a given m/z. However, even such constraints could still lead to ambiguous peak assignment possibilities, for example when overlapping ion signals at even higher m/z are encountered.

Over the course of more than 30 years, the analysis of mass spectra has evolved from simple ion peak position algorithms to complex spectral fitting routines. The importance of multiple peaks at identical unit masses, resulting in complex peak shapes was addressed as long ago as 1981 by van Katwijk, using mixtures of organic compounds from petroleum fractions [10], where multi-peak signals were due to isotopes and resolved by successively removing isotopes from ions at lower m/z. Feng et al. [11] used a Gaussian peak shape to determine the center of a peak in a mass spectrum recorded by a quadrupole mass spectrometer. A mass accuracy of 12 ppm was achieved. The situation of overlapping peaks was discussed as a problem for this type of peak fitting, but multi-peak fitting was not used to resolve multiple peaks. The asymmetry of peaks due to tuning issues of the mass spectrometer was addressed by Strittmatter et al. [12], where a multi-variate regression method and a double-Gaussian peak shape were used. The accuracy of 8 ppm from the peak fitting was a significant improvement over the 29 ppm of previously used centroid methods. The concept of baselines as an important component of mass spectra was discussed by Horn et al. [13], which determined baselines from histograms of the ion signals. House et al. [14] applied multi-peak fitting to overlapping features from a Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) mass spectrometer. That study showed the importance of using an instrument-specific mass resolution function to distinguish single wider peaks from multiple narrower peaks.

A complete procedure to fit multi-peak signals arising from the Aerodyne time-of-flight aerosol mass spectrometer (AMS) was described by DeCarlo et al. [15]. The algorithms developed in this study originate from that work. DeCarlo et al. [15] also used previously determined peak resolutions and shapes to constrain the multi-peak fitting method. A more recent study by Graus et al. [16] described how to identify multi-peak signals from proton-transfer reaction mass spectrometers (PTRMS) based on peak width considerations. They introduced an algorithm for assigning chemical formulas to individual high-resolution peaks based on a list of candidate formulas, constrained by bond valence rules. Subsequent studies [5,17,18] expanded on Graus et al. [15] and developed more refined methods and advanced algorithms for the PTRMS.

Here, a new analysis approach is described which can be used to assign the most likely ion formulas when the resolution of the mass spectrometer is not sufficient to unambiguously assign specific compound formulas. To our knowledge, there are no prior reports of techniques with similar goals in the literature. In this work, the mass spectrometer system was operated with a resolution of approximately 4000, which allows for some degree of separation of the many organic ion fragments but often not enough for unambiguous identification. The molecular and bulk level chemical information obtained with this analysis is demonstrated and evaluated with both field and simulated high-resolution spectra. A new method in which the raw mass spectrum is used to directly determine bulk chemical properties is also introduced. This method calculates bulk properties merely from signal intensity and mass defect without any peak assignment or fitting.

2. Methods

2.1. Experimental methods

All experiments used a high-resolution time-of-flight chemical ionization mass spectrometer, coupled to a micro-orifice volatility impactor inlet (MOVI-HRToF-CIMS, Aerodyne Research, Billerica, MA) [19,20]. The experimental data analyzed here were acquired during the BEACHON-ROMBAS field study in 2011 to provide organic composition information for gas and particle phase constituents. The measurement location was a pine forested area in the Colorado Rocky Mountains [21]. Details of the experimental setup are described elsewhere [22]. Briefly, ambient air was sampled using a custom inlet and aerosol particles were collected by impaction on a post while chemically analyzing the gas phase (henceforth referred to as "gas-phase cycle"). Then the particle composition was measured as the post was slowly heated leading to thermal desorption of the collected particles during the "particle cycle". Acetate (CH_3COO^-) was used as the reagent ion, which was produced by flowing acetic anhydride through a ^{210}Po ionizer [23] to selectively detect and quantify organic acids. As tuned, the mass spectrometer had a maximum resolution of 4000 and operated under "high declustering" conditions to avoid measuring ambient compounds clustered with reagent ions [24].

The methods developed here were applied to 15-s average spectra (from 1-s raw data) generated from the BEACHON-RoMBAS dataset (20–30 August, 2011). A "final peak list" was generated using highly averaged spectra (typically, 24-h averages, more than 80,000 mass spectra) to improve signal-to-noise (S/N) ratio and because some ions may not be present during all times of the day. It contained all possible peaks encountered from both gas- and particle-phase cycles. This list was eventually used with a strongly constrained fit algorithm to determine signal intensities for all measured raw spectra. The different components of the analysis steps are described below in detail. An overview of the data analysis flow is shown in Fig. 2.

2.2. Structure of the multi-peak fitting methods

All data analysis presented here was performed within the data analysis software *Igor Pro* (versions 6.2–6.3, Wavemetrics, Lake Oswego, OR, USA). The customized analysis routines are distributed by Tofwerk (Thun, Switzerland) and Aerodyne Research Inc. (Billerica, MA, USA) under the name "Tofware" (https://sites.google.com/site/citofms/analysis-software). The routines are designed for compatibility with Tofwerk TOF data files (HDF 5.0).