https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

Scaled Kendrick Mass Defect Analysis for Improved Visualization of Atmospheric Mass Spectral Data

Mitchell W. Alton[1,a], Harald Stark[1,2], Manjula R. Canagaratna[2], Eleanor C. Browne[1]

[1]Department of Chemistry and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado 80309, USA
[2]Aerodyne Research Inc., Billerica, Massachusetts, 01821, USA
[a]Now at the Aerosol Physics Research Group, University of Eastern Finland, Kuopio, 70211, Finland

Correspondence to: Eleanor C. Browne (Eleanor.Browne@Colorado.edu)

Abstract. Mass spectrometry is an important analytical technique within the field of atmospheric chemistry. Owing to advances in instrumentation, particularly with regards to mass resolving power and instrument response factors (sensitivities), hundreds of different mass-to-charge (*m/z*) signals are routinely measured. This large number of detected ions creates challenges for data visualization. Furthermore, assignment of chemical formulas to these ions is time-consuming and increases in difficulty at the higher *m/z* ranges. We present a technique called scaled Kendrick mass defect (SKMD) analysis to facilitate the visualization and peak identification processes for typical atmospheric organic (and to some extent inorganic) compounds. SKMD is related to the previously proposed resolution enhanced Kendrick mass defect (REKMD). SKMD introduces a tunable integer scaling factor into the mass defect equation that effectively contracts or expands the mass scale. The SKMD transformation maintains the horizontal alignment of ion series related by integer multiples of the chosen base unit that is characteristic of Kendrick mass defect analysis. However, the tunable integer acts to alter the mass defect spacing between different homologue ion series. As a result, the entire mass defect range (-0.5 to 0.5) is more effectively used simplifying data visualization and facilitating chemical formula assignment. We describe the mechanism of this transformation and discuss base unit and scaling factor selections appropriate for compounds typically found in atmospheric measurements. We present an open-source graphical user interface (GUI) for calculating and visualizing SKMD analysis results within the Igor Pro Environment.

1 Introduction

Recent improvements to the sensitivities, resolving power, and time-response of chemical ionization mass spectrometers used frequently in atmospheric measurements has led to a fundamental change in the understanding of atmospheric chemistry and the composition of the Earth's atmosphere. However, these advances have also created challenges in visualizing and interpreting the measurements. For typical resolving powers of time-of-flight mass spectrometers used in atmospheric chemistry, a conventional display of a mass spectrum as intensity versus mass-to-charge ratio *(m/z)* can only be

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

used to visually resolve the individual peaks across a narrow mass range. The intensity versus *m/z* visualization also gives little information about the composition of the ions being measured. To provide more chemical insight, various data visualization methods have been used to identify chemical relationships and trends. Some visualization methods display ions on a plot based on properties of their elemental composition, such as their H:C versus O:C ratios (van Krevelen plot; Van Krevelen, 1950) or average carbon oxidation state versus number of carbons (Kroll diagram; Kroll et al., 2011) of assigned ions. Other analyses relate compositional variables, such as the number of oxygen atoms, hydrogen atoms, or double bond equivalency of the assigned formula. However, the analyses just mentioned require formula assignments for each of the identified ions.

Analyses that do not rely on assigned chemical formulas of observed ions are advantageous for aiding in composition assignment and in visualizing data that contains ions of unassigned composition. One such analysis that can be visualized with minimal knowledge of the sample composition is plotting the mass defect, the difference between an ion's exact and nominal mass, against the nominal mass (Kendrick, 1963; Sleno, 2012) or exact mass. Since an ion's exact mass is determined by its elemental composition, an ion's mass defect retains compositional information. By plotting mass defect versus exact or nominal IUPAC mass, isobaric ions can be separated along the y-axis in a mass defect plot, thus improving the visualization (as compared to a typical intensity versus *m/z* mass spectrum) of closely spaced ions particularly across a wide mass range.

Kendrick mass analysis is one way in which mass defect analysis can be adapted to provide easier visualization of composition. In Kendrick mass analysis, the mass scale is redefined such that the mass of a base unit, R_{IUPAC}, is rounded to its integer value (Hughey et al., 2001; Kendrick, 1963). For our purposes, we assume singly charged ions and equate mass scale with mass-to-charge scale. Originally proposed using CH_2 as a base unit, Kendrick mass analysis has since been generalized to other base units (e.g., O, CH_2O, etc.). Equation 1 shows this transformation.

$$\mathrm{KM}(m/z, \mathrm{R_{IUPAC}}) = \frac{m}{z} \times \frac{\mathrm{round}(\mathrm{R_{IUPAC}})}{\mathrm{R_{IUPAC}}} \qquad (1)$$

where R_{IUPAC} is the IUPAC mass of the base unit R. The Kendrick mass defect is calculated using Equation 2.

$$\mathrm{KMD}(m/z, \mathrm{R_{IUPAC}}) = \mathrm{KM}(m/z, \mathrm{R_{IUPAC}}) - \mathrm{round}(\mathrm{KM}(m/z, \mathrm{R_{IUPAC}})) \qquad (2)$$

Note that the order of the terms in Equation 2 is determined mainly by convention within specific fields; we use the convention widely used in atmospheric chemistry. As a result of this transformation, ion series differing by an integer number of R units will have identical Kendrick mass defects. Typically, the result has been visualized in the two-dimensional space of Kendrick mass defect versus nominal Kendrick mass, however nominal IUPAC mass or exact IUPAC mass are also acceptable. In these spaces, homologous ion series differing by R will align horizontally. Traditional Kendrick mass defect analysis has proven to be an instrumental tool for visualizing mass spectral information from a variety of fields including petroleomics, proteomics, and atmospheric measurements (e.g., Junninen et al., 2010; Marshall and Rodgers, 2004; Sleno, 2012; Taguchi et al., 2010).

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

Kendrick mass analysis only requires the exact mass of the identified ion, not the assigned molecular formula, allowing for identification of ion series related by the molecular subunit R. Errors in the assignment of exact masses, particularly for ions with an unassigned elemental composition, will result in a "fuzzy" appearance to the horizontal alignment due to peak-fitting errors. Using traditional KMD analysis, the data points tend to only occupy a small fraction of the available KMD space (defined mathematically from -0.5 to +0.5) resulting in congested data visualizations that can make it challenging to identify homologous ion series. The limited range of the KMD arises because of "dead-space" between the masses of common chemical formulas. Particularly for compounds present in complex environmental mixtures, observed ions masses tend to be periodically spaced with ~1 atomic mass unit (amu) gaps and the Kendrick transformation maintains this spacing. The existence of the dead-space can be explained because environmental molecules are generally made of a limited number of elements (H, C, O, N, S).

Recently Fouquet and Sato (Fouquet et al., 2018; Fouquet and Sato, 2017c, 2017a, 2017b) have introduced the concept of resolution enhanced Kendrick mass defect (REKMD) analysis to provide improved visualization and analysis of mass spectrometry data, particularly for polymers. REKMD introduces the concept of fractional base units by using integer divisors (X) as shown in Equation 3:

$$\mathrm{REKM}(m/z, \mathrm{R_{IUPAC}}, X) = \frac{m}{z} \times \frac{\mathrm{round}\left(\frac{\mathrm{R_{IUPAC}}}{X}\right)}{\frac{\mathrm{R_{IUPAC}}}{X}} \quad (3)$$

REKMD is defined analogously to Equation 2 using REKM rather than KM, as shown in Equation 4.

$$\mathrm{REKMD}(m/z, \mathrm{R_{IUPAC}}, X) = \mathrm{REKM}(m/z, \mathrm{R_{IUPAC}}, X) - \mathrm{round}(\mathrm{REKM}(m/z, \mathrm{R_{IUPAC}}, X)) \quad (4)$$

For integer values of X, ions differing by integer numbers of R will have identical REKMD values. Specific non-integer values of X can also be used as shown previously (Fouquet and Sato, 2017b). Appropriate selection of X amplifies mass defect variations increasing the range of mass defect space occupied by a given dataset and improving horizontal alignment of homologous ion series. REKMD analysis method has been used in polymer chemistry previously (e.g., Fouquet et al., 2018; Fouquet and Sato, 2017a, 2017b), but to our knowledge has not been previously applied to atmospheric samples. It should be emphasized that the REKM/REKMD transformation has no impact on the mass resolution of the data, but rather alters the separation of ions in mass-defect space. Through appropriate selection of X, the separation in mass defect space can be tuned to enable easier visualization of homologous ion series resulting in an apparent "resolution enhancement."

In this work, we present a variation of REKMD analysis, which we term scaled Kendrick mass defect (SKMD) and discuss in general terms the principles of the mechanisms by which the mass defect space is expanded. We demonstrate its application for visualization of atmospheric trace gas composition, describe how choices of R and X, which we term scaling factor when used in SKMD, will affect the visualization, show how the technique can aid in molecular formula assignment to unknown ions, and describe an open-source graphical user interface (GUI) for performing the analysis. We suggest that this analysis can be used not only for understanding ambient atmospheric gas-phase measurements as shown here, but could have potential use in aerosol measurements, and more broadly for other types of mass spectrometric data.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

2 Vocus Proton Transfer Mass Spectrometer

For illustrating the applications of SKMD analysis for atmospheric chemistry, we use measurements from an Aerodyne and Tofwerk Vocus Proton-transfer mass spectrometer. Details of this instrument are discussed elsewhere (Krechmer et al., 2018). This measurement technique is commonly used in atmospheric chemistry as it can detect and quantify a large number of hydrocarbons (with the exception of small alkanes) as well as oxygen, nitrogen, and sulfur containing organic molecules found in the environment (Sekimoto et al., 2017). The instrument was deployed in Billerica, MA from March to August of 2021, with 1 Hz data averaged to 30 minutes before analysis. For the purposes of this discussion, we will be discussing the data collected on July 9, 2020, from 4:00 to 23:00 local time (UTC – 4). All data was analyzed in Tofware v3.2.5 within the Igor Pro v9.0.0.10 environment (Wavemetrics, Inc., Portland, OR). Only signals above a certain threshold (1 count per second) and which changed more than 30% between evening and morning were included in the analysis. The reagent ions were also removed from the analysis. This ambient dataset is used in Sect. 3 to demonstrate the principles of SKMD and illustrate how different scaling factors separate mass spectral data. The same data is used in Sect. 4 to present how SKMD analysis can aid in understanding chemical composition in measured mass spectra. Individual ion signals are also purposefully unassigned and refit to demonstrate the usefulness of this tool for determining unidentified signals.

3 SKMD – Concepts and Method

3.1 Scaled Kendrick Mass Defect

Traditional Kendrick mass defect analysis uses round(R_{IUPAC}) as the nominal mass for the mass scale transformation (Equation 1); however, it is mathematically acceptable to use other integer values to maintain horizontal alignment of ion series related by an integer number of R. In fact, one can expand or contract the mass scale by replacing round(R_{IUPAC}) in Equation 1 with an integer scaling factor X as in Equations 5 and 6:

$$\mathrm{SKM}(m/z, \mathrm{R_{IUPAC}}, X) = \frac{m}{z} \times \frac{X}{\mathrm{R_{IUPAC}}} \quad (5)$$

$$\mathrm{SKMD}(m/z, \mathrm{R_{IUPAC}}, X) = \mathrm{SKM}(m/z, \mathrm{R_{IUPAC}}, X) - \mathrm{round}(\mathrm{SKM}(m/z, \mathrm{R_{IUPAC}}, X)) \quad (6)$$

with X values less than round(R_{IUPAC}) contracting the scale and values greater than round(R_{IUPAC}) expanding the scale. When analyzing mass spectral data with Kendrick mass defect, the main goal is to identify horizontal lines of ions related by integer numbers of R. When X is introduced into the equation and the scale changes, this horizontal alignment is preserved, however, the lines are separated more clearly in the mass defect dimension allowing for simpler identification of related ions. We label this scaled Kendrick mass (SKM) and scaled Kendrick mass defect (SKMD). For the two-dimensional visualizations of scaled Kendrick mass defect versus mass, we find the exact or nominal IUPAC mass rather than SKM to be the most intuitive x-axis.

SKMD (Eq. 6) is mathematically identical to REKMD (Equation 4) for integer scaling factors (X) satisfying Equation 7:

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

$$\text{round}\left(\frac{2 \times \text{R}_{\text{IUPAC}}}{3}\right) < X \leq \text{round}(2 \times \text{R}_{\text{IUPAC}}) \quad (7)$$

since round(R_{IUPAC}/X) will equal 1. This range of X coincides with the recommended range of integer divisors for REKMD analysis (Nakamura et al., 2019). SKMD differs from REKMD in that the mass defect expansion is linear in X at values of $X \leq$ round(2 R_{IUPAC}/3) unlike the non-linear expansion for REKMD (Fig. S1). These smaller X values can be useful when analyzing large mass ranges. Additionally, unlike REKMD, no upper limit on X exists for SKMD analysis (Fig. S1) which can be useful in tuning the separation of homologous ion series when larger separation in the y-axis is desired as will be discussed later. However, unlike REKMD for which select real values of X provide mathematically acceptable solutions, only integer X are allowed for SKMD to maintain horizontal alignments. Integer values provide similar separation in REKMD analysis as real values (Fouquet et al., 2019), suggesting that limiting X to integers does not restrict this analysis.

Compared to a traditional KM/KMD analysis, SKM/SKMD has two advantages: 1) for select combinations of R and X, SKMD visualization will provide increased information on chemical composition and 2) the increased range of mass defect space will lead to clearer alignment of homologue ion series enabling better visual identification and potentially aiding in chemical formula assignment of ions. For both KMD and SKMD analysis, ions differing by integer units of R will align horizontally in these spaces. Note that any R can be used, though for the purposes of this work we focus on the divisors when the Kendrick base is ^{16}O.

3.2 Visualization of chemical composition

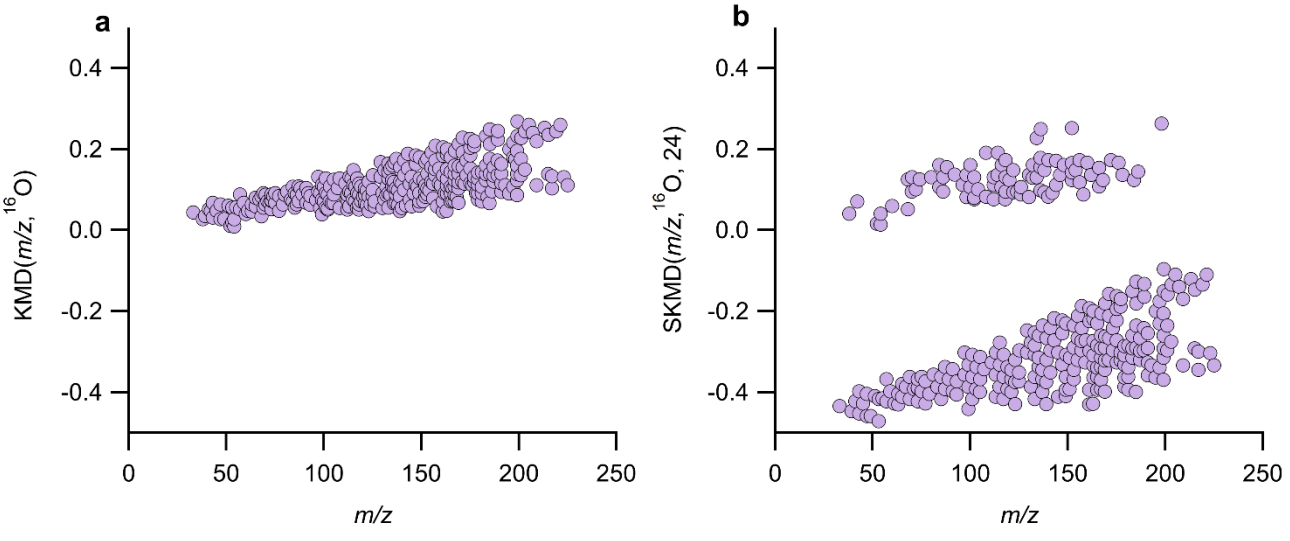


Figure 1 Using ambient data collected by the Vocus in Billerica, MA, (a) traditional KMD plot using a base of ^{16}O and (b) SKMD plot using a base of ^{16}O and X=24, where the two groupings correspond to even (positive SKMD values, odd number of nitrogen atoms) and odd (negative SKMD values, zero/even number of nitrogen atoms) nominal m/z. Fig. S2 shows (a) zoomed in to illustrate that ion alignment remains blurred even with different y-axis scaling.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

The combined choice of R_{IUPAC} and X impacts the mass scale expansion/contraction and will dictate how SKMD analysis aids visualization of composition and alignment of homologous ion series. Figs. 1a and 1b compare KMD(*m/z*, ^{16}O) and SKMD(*m/z*, ^{16}O, 24) . For the SKMD(*m/z*, ^{16}O, 24) analysis, the mass scale is expanded by a factor of ~3/2. As a result of this scaling, ions with odd nominal masses in IUPAC mass space will be shifted towards half-integer masses while even nominal masses in IUPAC space will remain at approximately integer values. Assuming positive mass defects in IUPAC mass space, nominally odd mass ions will typically have negative scaled Kendrick mass defects and nominally even mass ions will have positive scaled Kendrick mass defects leading to the two groupings in Fig. 1b. This transformation of the scaled Kendrick masses is also shown in Fig. 2. Figs. 2a and 2b show how ions in IUPAC *m/z* or KM space span a narrow mass defect range whereas Fig. 2c shows that for the SKM(*m/z*, ^{16}O,24), a transformation of ~3/2, nominally odd mass ions (in IUPAC *m/z*) have a SKMD(*m/z*, ^{16}O, 24) of around -0.5 while the SKMD(*m/z*, ^{16}O,24) of even mass ions remains around zero.

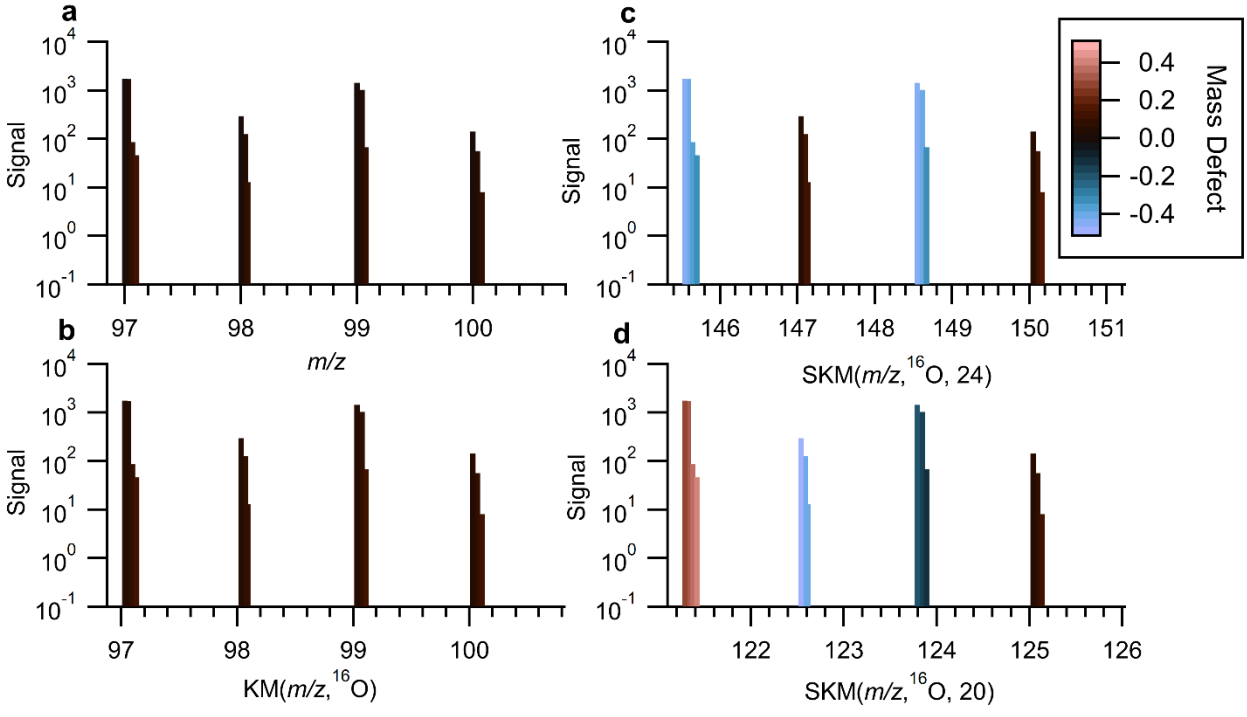


Figure 2 (a) An example section of a mass spectra of ambient data measurements plotted against the IUPAC mass-to-charge values. (b) The same series of identified peaks plotted against Kendrick mass with a base unit of ^{16}O, (c) scaled Kendrick mass with R=^{16}O and *X*=20, and (d) scaled Kendrick mass with R=^{16}O and *X*=24. In all plots, the identified ions are colored by the mass defects after the mass transformation.

By separating even and odd nominal IUPAC masses into different regions, the SKMD(*m/z*, ^{16}O, 24) visualization provides information on chemical composition not available with a standard KMD plot. Specifically, for de-isotoped data sets comprised of compounds following the nitrogen rule, the two groups will represent compounds with odd or even/zero

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

nitrogen atoms. For atmospheric chemistry measurements, compounds with two or more nitrogen atoms are usually minor (both in abundance and in number of species) compared to compounds with no nitrogen atoms and thus the SKMD with X/R_{IUPAC} of ~3/2 provides visual information on nitrogen versus non-nitrogen containing compounds. A notable exception would be situations in which organic dinitrates are abundant. In our data, we identified only 15 compounds (3% of the total number of ions included in analysis) that contained 2 nitrogen atoms. Although other methods can be used to separate even and odd *m/z* (masking, making multiple plots, etc.), SKMD can separate the even and odd masses on the same plot making comparison between the groups of ions simpler.

One can intuit how the choice of R and X affects the degree of expansion through inspection of the approximate reduced fraction of X/R_{IUPAC}. The reciprocal of the denominator of the reduced fraction represents the fractional mass intervals IUPAC integer masses are transformed to. As such, the number of groupings from a certain transformation is the reciprocal of the denominator. For instance, for R=^{16}O and X = 8 or 24, the approximate fractions are 1/2 and 3/2 and thus interval IUPAC masses will be transformed to half-integer and integer SKM masses (Fig. 2c) resulting in two groupings. Fig. S3 shows the results of X/R_{IUPAC} of ~3/2 for other choices of R_{IUPAC}. For R = ^{16}O and X = 4, 12, or 20, the approximate reduced fractions are 1/4, 3/4, and 5/4 respectively and all these choices will transform even IUPAC integer masses to integer or half-integer SKM values and odd IUPAC integer masses to quarter and three-quarter integer values (Fig. 2d). Thus, SKMD values will roughly start around 0.0, ±0.25, and ±0.5 and this transformation result in four "groupings" of SKMD values (Fig. 3a). Although four groupings will result for X = 4, 12, or 20, the exact SKMD value of a given ion will depend on X. Likewise, for R=^{16}O and X = 2, 6, 10, or 14 will result in 8 groupings with the groups representing alternating even and odd nominal IUPAC masses. When the denominator of the reduced fraction is large, as would happen for X =17 with R=^{16}O (a reduced fraction of 17/16), the groupings overlap significantly (Fig. 3b). For odd denominators, such as encountered for R=$^{12}CH_2$ with X = 8 (approximate reduced fraction of 4/7), the groups will no longer correspond to even/odd nominal IUPAC masses, but rather a different metric, and thus the SKMD visualizations will provide alternate but complimentary information. Despite the limitations in these last two examples, visualization can still be improved compared to a traditional KMD plot since the homologous series will be separated more clearly into individual horizontal lines, as seen in Fig. 3b. These other scaling factors may be useful when looking at spectra with fewer identified ions, as separating horizontal homologous ion series can be more useful than creating groupings of ions with the same number of nitrogen atoms or other grouping criteria.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

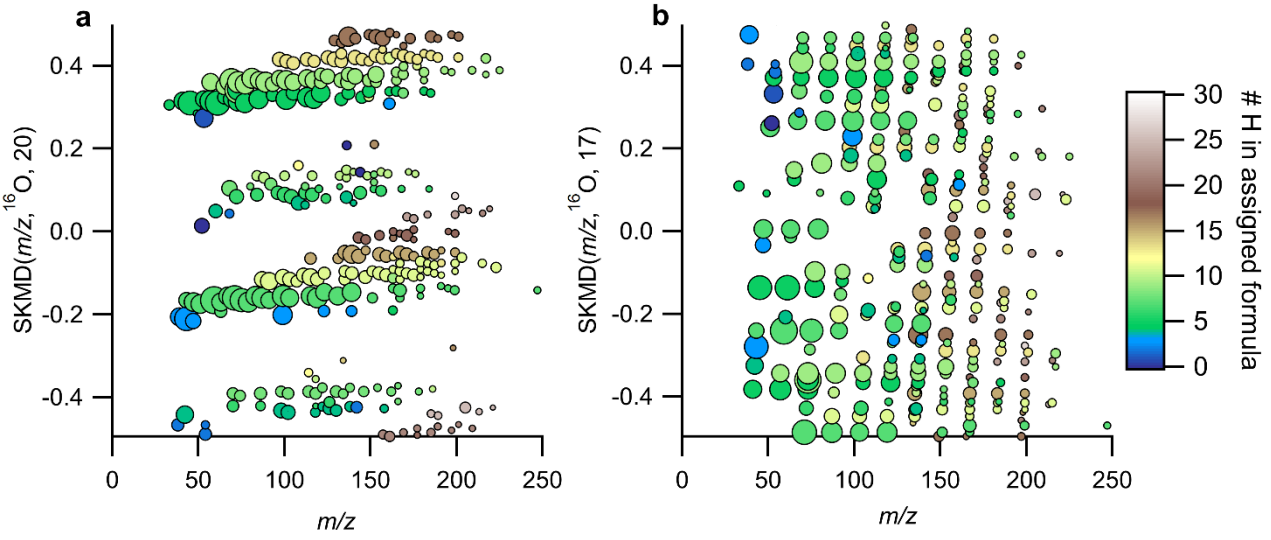


Figure 3 (a) SKMD plot of data obtained from Vocus ambient measurements with base of ^{16}O, and (a) $X = 20$ (b) $X = 17$. The points are colored by the number of hydrogens in the assigned formula and sized by the log of the measured intensity.

The numerator of the reduced fraction is important for understanding the degree of expansion/contraction of the mass scale. At low numerator values, the mass scale contraction reduces the spread of SKMD values around a given nominal IUPAC mass, while higher numerator values increase the spread (Fig. 4). At sufficiently high values of *X*, "aliasing" or "wrap-around" is introduced (+0.5 is transformed to -0.5), which can be seen in Fig. S4 when *X*=20 around *m/z* 100-250. Aliasing is non-linear with *X* and is more common when dealing with divisors that give increased numbers of "groupings" thus explaining why *X*=40 (approximate reduced fraction of 5/2) displays negligible aliasing compared to *X*=20 (approximate reduced fraction of 5/4; Figs. 4 and S4) As aliasing can complicate the interpretation of the data, it is recommended to either manually anti-alias data (most applicable for small data sets) or select *X* that maximally expands the data in SMKD space while also minimizing aliasing. This can be determined by plotting the defect spreads (difference between highest and lowest SKMD(*m/z*, R_{IUPAC}, X)) as a function of *m/z* with various values for *X*.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

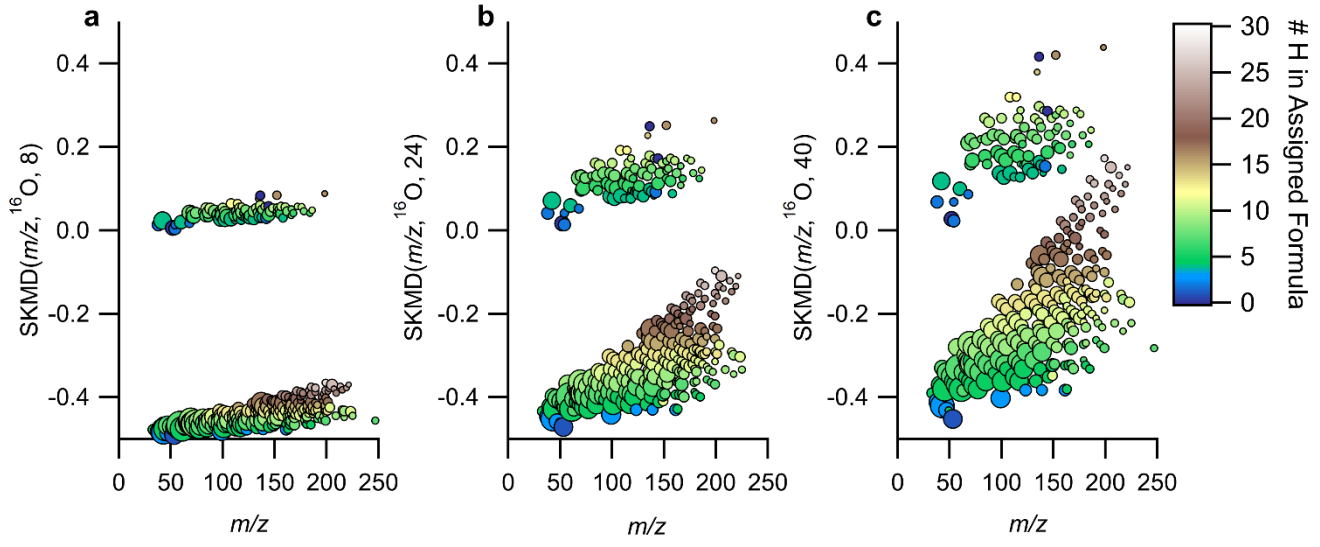


Figure 4 Expansion of the same data in SKMD space as the values for *X* are increased while using ^{16}O as R. An *X* value of 40 would not be possible in REKMD analysis, however, for the data visualized here, it provides acceptable resolution. The points are colored by the number of hydrogens in the assigned formula and sized by the log of the measured intensity.

In addition to the even/odd nominal IUPAC *m/z* separation, and corresponding information on the number of nitrogen atoms discussed earlier, select combinations of *X* and R_{IUPAC} provide further information on chemical composition. For instance, for $C_xH_yO_zN_w$ compounds for w of 0 and/or 1 and base units of R_{IUPAC} ^{16}O or ^{12}C, select values of *X* will lead to grouping of compounds with the same number of hydrogen atoms in the same area of the SKMD plot (e.g., Figs. 3a and 4). Moreover, within each grouping of a constant number of hydrogen atoms, each horizontal line will correspond to a constant number of carbon atoms when using a base of ^{16}O or a constant number of oxygen atoms when using a base of ^{12}C. For a base unit of ^{16}O, the number of carbon atoms will increase as one moves towards more positive mass defects while for the base unit of ^{12}C, the number of oxygen atoms will increase as one moves towards more negative mass defects. The separation by number of hydrogen atoms (and other groupings) is further explained in Sect. S2 and Fig. S5 of the Supplement.

As in traditional KMD analysis, select choices of R_{IUPAC} provide information on double bond equivalency (DBE), an estimation of the number of double bonds (or degrees of unsaturation, including rings) in an elemental formula shown in Equation 8:

$$\mathrm{DBE} = \mathrm{c} - \frac{\mathrm{h}}{2} + \frac{\mathrm{n}}{2} + 1 \qquad (8)$$

where c, h, and n are the number of carbon, hydrogen, and nitrogen atoms in the formula, respectively. For R_{IUPAC} of ^{16}O or $^{12}CH_2$ horizontal lines correspond to constant DBE while for ^{12}C, DBE will increase moving from left to right across a horizontal line.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

3.3 Improved visual alignment of homologue ion series

Figs. 5a and 5b show an example of the improved visual alignment of homologue ion series. Both panels contain the same number of points, with ~50% of the points identical in both panels. The horizontal alignment of the points is visually clearer with SKMD(*m/z*, ^{16}O, 24) compared to normal KMD(*m/z*, ^{16}O). The apparent improvement in alignment results from the increased vertical spacing between the different horizontal lines. This increase in spacing is achieved by increasing the mass defect range occupied by the data and by moving the masses at +1 *m/z* to a different area of the SKMD(*m/z*, ^{16}O, 24) plot. Once the identified ions are separated into related groupings, using the software tool presented here, a subset of these ions can be easily selected and re-analyzed with a different R_{IUPAC} and *X*, as will be discussed in Sect. 3.4, providing more in-depth information about a specific subset of ions.

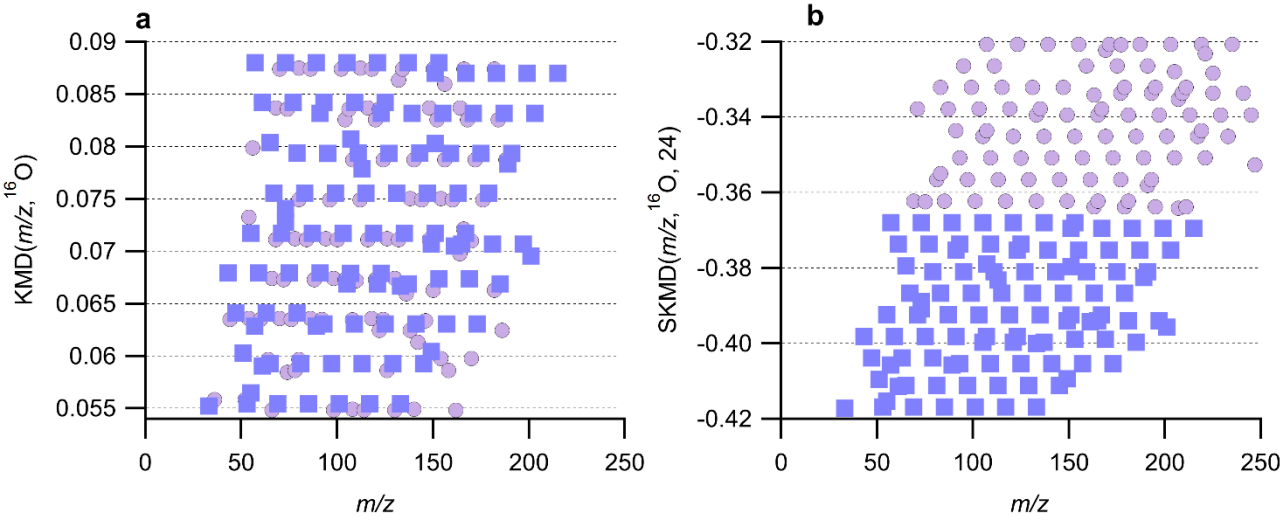


Figure 5 A zoomed in section of the (a) KMD plot from Fig. 1a and (b) the SKMD plot from Fig. 1b. Both subpanels contain the same number of total points. The square purple points correspond to the same ions in the subpanels. The circle lilac points represent ions unique to each subpanel. Note the different y-axis range in each subpanel.

3.4 GUI for REKMD Analysis in Igor Pro Environment

The data in this work was analyzed using a graphical user interface (GUI) we built that operates inside the Igor Pro Environment (Wavemetrics, Lake Oswego, OR; Igor Pro v9 and above). The GUI allows the user to select a data set to perform SKMD analysis with the R and *X* of their choice. The code currently has $^{12}CH_2$, ^{16}O, ^{14}N, ^{12}C, and isoprene (C_5H_8) available to choose from, though other bases can be added to the list by small modifications to the code. The GUI also provides optional inputs for intensity data for sizing/coloring of the points made in the SKMD plots. The GUI allows interactive point filtering by providing an option for the user to draw a polygon around a set of points and recalculate the SKMD plot on just those points, with the option of using a different R or *X* for the analysis. Filtering options are included to

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

remove the points with the largest and smallest signals for easier visualization. The code for the GUI is available for download from GitHub, with more information in Sect. S3 and Fig. S6 of the Supporting Information.

4 Example Applications of REKMD Analysis

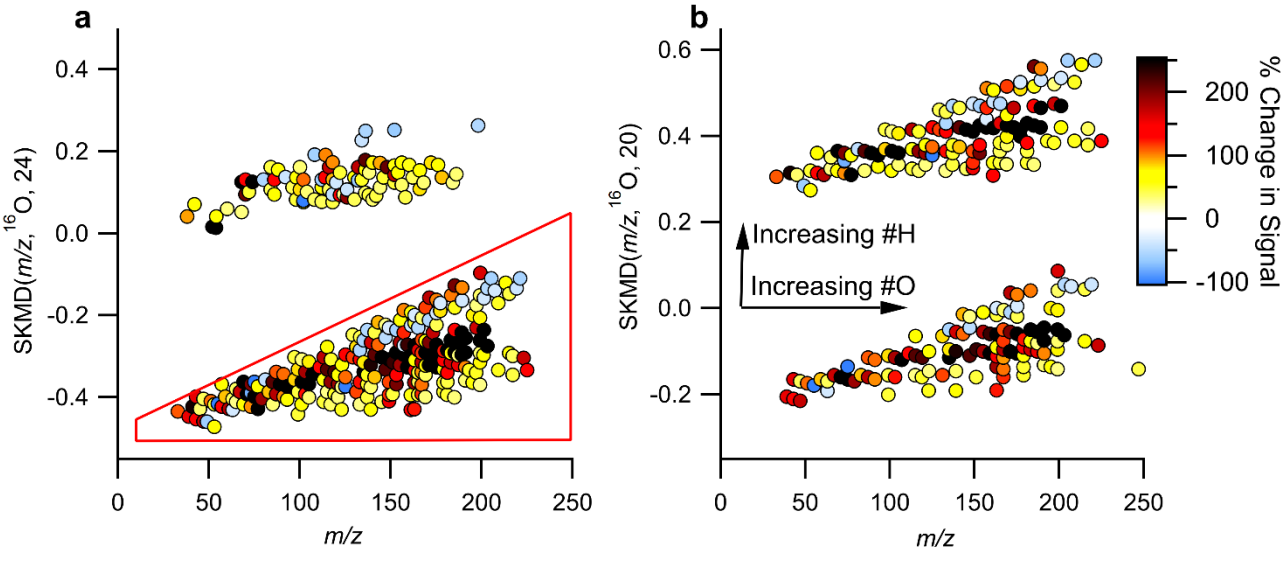
4.1 Visualizing Composition

To explore the utility of SKMD, we delve further into the data collected in Billerica, MA presented in the previous sections to show how homologous ion series alignments can be used. Ions were assigned based on high-resolution fully constrained peak fitting (Cubison and Jimenez, 2015; Stark et al., 2015) though this analysis technique works without prior knowledge of the molecular formulas, just the exact measured *m/z*. High-resolution, fully constrained, peak fitting can possibly aid in determining if a peak assignment is missing from the measured data; however, this becomes more difficult at higher *m/z* and with higher complexity samples (Cubison and Jimenez, 2015; Timonen et al., 2016). An example of the peak fitting has been demonstrated previously (Cubison and Jimenez, 2015). In Fig. 6, the SKMD plots of the ambient data collected in Billerica, MA are shown, with points colored by the percent change in intensity of the signal between morning and evening. For this analysis, we focus on early morning and late afternoon as times when emissions, photochemistry, and dynamics are known to be different. The ions plotted are limited to those which satisfy the following conditions: 1) have an average intensity above 1 count per second (cps), 2) change more than 30% between the morning and evening, 3) are not the primary reagent ions. The percent difference is calculated as the difference between the morning and evening integrated intensities divided by the intensity in the morning, leading to positive values reflecting an increase in signal in the afternoon compared to the morning. As some ions have intensities of 0 ions/s in the morning, the percent change can be undefined, therefore points with percentage increases greater than 250%, including undefined increases, are the same color.

Fig. 6a shows that the points at the center of each of the groups increase the most (colored black), while those with slightly higher or lower SKMD(*m/z*, ^{16}O, 24) increase less or even decrease. Additionally, as this divisor separates odd and even *m/z*, this plot also shows that the odd *m/z* ions, consisting of $(C_xH_yO_z)H^+$ compounds and compounds with an even number of nitrogen atoms (assuming closed electron shell molecules ionized via proton transfer), have the largest fractional increase. Using the polygon selection tool in the GUI, we can reperform the SKMD analysis on just the $C_xH_yO_z$ compounds (and the 15 identified $C_xH_yO_zN_{2w}$ compounds). Fig. 6b shows the results of performing the SKMD analysis on this subset of data using a different *X*, in this case, 20. Note that manual anti-aliasing has been applied. With an approximate reduced fraction of 5/4, this new transformation would nominally result in 4 groupings, however since only ions with odd nominal IUPAC *m/z* were included, only 2 groupings are visible. These groupings are separated by ~2 amu in IUPAC *m/z* space and, as such, chemical formulas will be related by the addition of 2 hydrogen atoms. For instance, $C_7H_{10}O_5H$+ will be in the lower group while $C_7H_{12}O_5H$+ will be in the higher group with SKMD(*m/z*, ^{16}O, 20) values of -0.105 and 0.415 respectively Arrows are included to show the transitions between chemical formulas within the groupings. This plot has the advantage of further

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

280 spacing out the ions allowing for clearer chemical groupings and alignments. It shows that the compounds with the greatest number of hydrogen atoms (those at the top of each grouping) decrease the most between the morning and evening hours. Some of the signals that increase the most have 5, 9, or 10 carbon atoms, suggesting they could be from isoprene or monoterpene oxidation over the course of the day. Some specific formulas (and potential identifications) that increase are $(C_5H_8)H^+$ (isoprene or an isomer or an ion fragment), $(C_5H_{10}O_4)H^+$ (a monosaccharide), $(C_9H_{14}O_4)H^+$, and $(C_{10}H_{17}O_4)H^+$
285 (possible monoterpene oxidation products). These ion signals could correspond to the emission and oxidation of biogenic compounds, such as terpenes, which are anticipated to increase as biological activity and atmospheric oxidation occurs. Relatedly, the compounds that increased the most have either 9, 11, 13, or 15 hydrogen atoms in the assigned formulas, including the proton from ionization. This analysis can aid in understanding general atmospheric chemistry and how oxidation affects molecular structures and saturation in a bulk method.



290

Figure 6 (a) SKMD plot using X=24 with a ^{16}O base. Points are colored by the percent change in the signal between the morning and evening during one day of measurements. (b) Using the points selected with the polygon tool in the GUI, the SKMD plot is remade using X=20. The arrows correspond to the changes in an individual grouping. Another split is created when re-calculating SKMD with a different X based on number of hydrogen atoms in the formulas. Note that manual anti-aliasing has been applied in panel (b) to keep
295 related ions together and that the y-axis range of the two subpanels differs.

4.2 Using REKMD for Chemical Formula Assignment

By increasing the separation in mass defect space, SKMD analysis can aid in chemical formula assignment, particularly when extension of homologue series is an appropriate tool for aiding in assignment. SKMD analysis can also provide insight into potential ion misassignment. Fig. 7a shows a normal KMD plot, KMD(*m/z*, ^{16}O), while Fig. 7b shows SKMD(*m/z*, ^{16}O,
300 20). The colored points are the same identified ion signals in both figures. The points are colored by the number of carbon atoms and all formulas have 9 hydrogen atoms in the assigned formula. The turquoise points are assigned ions that were

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

removed, then added back into the peak list as "unknowns" with Tofware's automatic peak fitting procedure. The grey points in Fig. 7a are points that are not present in Fig. 7b. This figure shows that there is significantly more overlap with other ions in a KMD(*m/z*, ^{16}O, 16) plot than SKMD(*m/z*, ^{16}O, 20). The separation of ions can aid in ion formula identification. As horizontal lines are made for ions with increasing number of ^{16}O atoms in the formula, and the same number of hydrogens and carbon atoms, identifying missing ions is a simple matter of adding or subtracting an O atom from the adjacent formula to find the missing ion. These horizontal relationships can be useful for automatic or semi-automatic ion identification in the future. Note that the homologous ion series are still present in Fig. 7a but are just visually more difficult to see without prior knowledge of the ion identities.

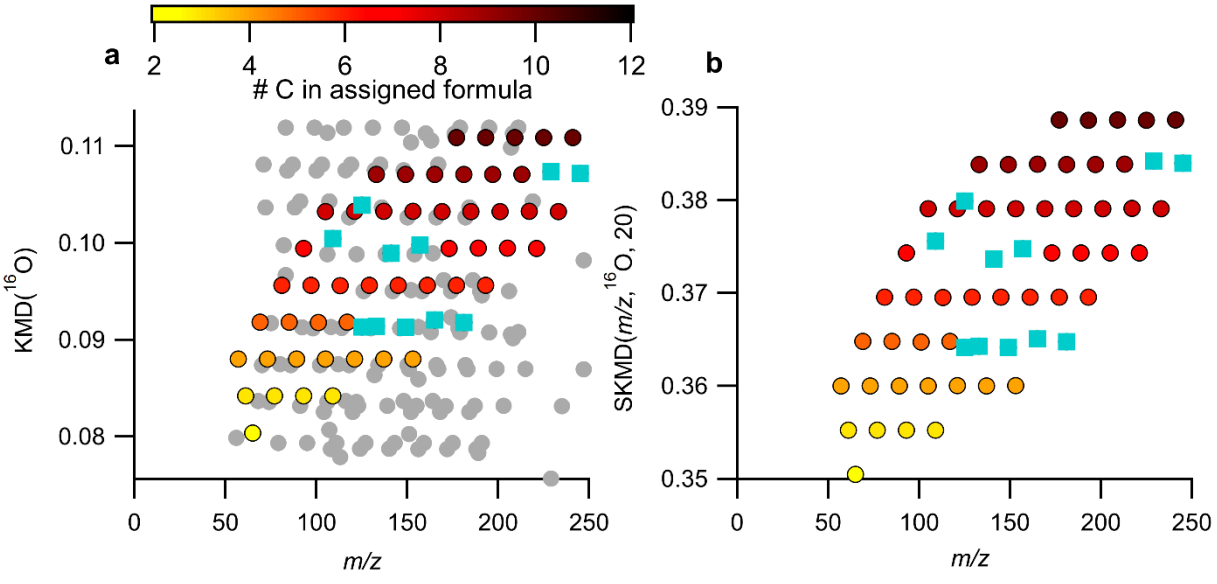


Figure 7 A subset of the ambient VOCUS data in (a) **a** KMD plot with base ^{16}O and (b) a SKMD plot using ^{16}O as the base and *X*=20. The grey points in (a) are points that appear within the field of view due to the compression of the KMD space. Colored points are the same ions in both subpanels. Circles are colored by the number of carbon atoms in the formula. Ions that were removed and then re-added using Tofware's built-in peak addition, with no human intervention to improve the fitting, are shown in blue squares. Note that in order to better represent the peak assignment process, this figure contains all identified peaks, not just those passing the filtering criteria identified in Sects. 2 and 4.1.

5. Conclusions

We present SKMD analysis as a technique to improve visualization and peak identification in mass spectrometric measurements, particularly for atmospheric measurements. As demonstrated here, this method can aid in the identification of unknown ions and show chemical trends in a clearer manner than regular Kendrick mass defect plots. Additionally, with appropriate selection of *X*, certain classes of ions can be grouped, such as by the number of hydrogen atoms or the number of nitrogen atoms. Using this separation technique, ions can be more easily characterized and visualized, allowing for easier interpretations and assignments of chemical formulas. This analysis can be used as an initial tool to better understand what

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

ions change more over the course of a measurement, identify which ions are likely misidentified, and facilitate interpretation of the measured chemical composition. We focus on gas-phase atmospheric measurements for our analysis, but SKMD can be applied to the mass spectra obtained from aerosol samples, with promise as a tool to understand polymerization products' contributions to aerosol. Additionally, these plots will be beneficial for something like looking at the chemical composition of positive matrix factorization (PMF) factors.

Code Availability. Procedure file containing the code to run the SKMD panel in the Igor Pro v9 environment (.ipf) is available at https://github.com/BrowneLab/SKMD_Panel.git

Supplement. The supplement related to this article is available online.

Competing interests. The authors declare they have no competing interests.

Present Address. †Now at the Aerosol Physics Research Group, University of Eastern Finland, Kuopio, 70010

Author Contributions. ECB conceptualized the work. MWA and ECB developed SKMD and the applications to atmospheric chemistry data sets. Data was collected by MC. Code was developed by MWA with guidance from ECB and HS. All authors have given approval to the final version of the manuscript. The manuscript was written by MWA with guidance from ECB.

Acknowledgments. MWA was supported by the Cooperative Institute for Research in Environmental Sciences Graduate Research Award during this work.

References

Cubison, M. J. and Jimenez, J. L.: Statistical precision of the intensities retrieved from constrained fitting of overlapping peaks in high-resolution mass spectra, Atmos. Meas. Tech., 8(6), 2333–2345, doi:10.5194/amt-8-2333-2015, 2015.

Fouquet, T. and Sato, H.: Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units, Anal. Chem., 89(5), 2682–2686, doi:10.1021/acs.analchem.6b05136, 2017a.

Fouquet, T. and Sato, H.: How to choose the best fractional base unit for a high-resolution Kendrick mass defect analysis of polymer ions, Rapid Commun. Mass Spectrom., 31(12), 1067–1072, doi:10.1002/rcm.7868, 2017b.

Fouquet, T. and Sato, H.: Improving the Resolution of Kendrick Mass Defect Analysis for Polymer Ions with Fractional Base Units, Mass Spectrom., 6(1), A0055–A0055, doi:10.5702/massspectrometry.A0055, 2017c.

Fouquet, T., Satoh, T. and Sato, H.: First Gut Instincts Are Always Right: The Resolution Required for a Mass Defect Analysis of Polymer Ions Can Be as Low as Oligomeric, Anal. Chem., 90(4), 2404–2408, doi:10.1021/acs.analchem.7b04518, 2018.

Fouquet, T. N. J., Nakamura, S., Sato, H. and Cody, R. B.: Real divisors and pseudo-continuous enhancement of resolution for a Kendrick mass defect analysis, Rapid Commun. Mass Spectrom., 33(19), 1547–1551, doi:10.1002/rcm.8500, 2019.

Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G. and Qian, K.: Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra, Anal. Chem., 73(19), 4676–4681, doi:10.1021/ac010560w, 2001.

https://doi.org/10.5194/egusphere-2022-1319
Preprint. Discussion started: 9 December 2022

Junninen, H., Ehn, M., Petäjä, Luosujärvi, L., Kotiaho, T., Kostiainen, R., Rohner, U., Gonin, M., Fuhrer, K., Kulmala, M. and Worsnop, D. R.: A high-resolution mass spectrometer to measure atmospheric ion composition, Atmos. Meas. Tech., 3(4), 1039–1053, doi:10.5194/amt-3-1039-2010, 2010.

fKendrick, E.: A Mass Scale Based on CH 2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds., Anal. Chem., 35(13), 2146–2154, doi:10.1021/ac60206a048, 1963.

Krechmer, J., Lopez-Hilfiker, F., Koss, A., Hutterli, M., Stoermer, C., Deming, B., Kimmel, J., Warneke, C., Holzinger, R., Jayne, J., Worsnop, D., Fuhrer, K., Gonin, M. and De Gouw, J.: Evaluation of a New Reagent-Ion Source and Focusing Ion-Molecule Reactor for Use in Proton-Transfer-Reaction Mass Spectrometry, Anal. Chem., 90(20), 12011–12018, doi:10.1021/acs.analchem.8b02641, 2018.

Van Krevelen, D.: Graphical statistical method for the study of structure and reaction processes of coal, Fuel, 29(269), 1950.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E. and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nat. Chem., 3(2), 133–139, doi:10.1038/nchem.948, 2011.

Marshall, A. G. and Rodgers, R. P.: Petroleomics: The Next Grand Challenge for Chemical Analysis, Acc. Chem. Res., 37(1), 53–59, doi:10.1021/ar020177t, 2004.

Nakamura, S., Cody, R. B., Sato, H. and Fouquet, T.: Graphical Ranking of Divisors to Get the Most out of a Resolution-Enhanced Kendrick Mass Defect Plot, Anal. Chem., 91(3), 2004–2012, doi:10.1021/acs.analchem.8b04371, 2019.

Sekimoto, K., Li, S.-M., Yuan, B., Koss, A., Coggon, M., Warneke, C. and de Gouw, J.: Calculation of the sensitivity of proton-transfer-reaction mass spectrometry (PTR-MS) for organic trace gases using molecular properties, Int. J. Mass Spectrom., 421, 71–94, doi:10.1016/j.ijms.2017.04.006, 2017.

Sleno, L.: The use of mass defect in modern mass spectrometry, J. Mass Spectrom., 47(2), 226–236, doi:10.1002/jms.2953, 2012.

Stark, H., Yatavelli, R. L. N., Thompson, S. L., Kimmel, J. R., Cubison, M. J., Chhabra, P. S., Canagaratna, M. R., Jayne, J. T., Worsnop, D. R. and Jimenez, J. L.: Methods to extract molecular and bulk chemical information from series of complex mass spectra with limited mass resolution, Int. J. Mass Spectrom., 389, 26–38, doi:10.1016/j.ijms.2015.08.011, 2015.

Taguchi, V. Y., Nieckarz, R. J., Clement, R. E., Krolik, S. and Williams, R.: Dioxin analysis by gas chromatography-fourier transform ion cyclotron resonance mass spectrometry (GC-FTICRMS), J. Am. Soc. Mass Spectrom., 21(11), 1918–1921, doi:10.1016/j.jasms.2010.07.010, 2010.

Timonen, H., Cubison, M., Aurela, M., Brus, D., Lihavainen, H., Hillamo, R., Canagaratna, M., Nekat, B., Weller, R., Worsnop, D. and Saarikoski, S.: Applications and limitations of constrained high-resolution peak fitting on low resolving power mass spectra from the ToF-ACSM, Atmos. Meas. Tech., 9(7), 3263–3281, doi:10.5194/amt-9-3263-2016, 2016.