

We thank the referee's and the public commentor for their constructive feedback. We have made several changes to the manuscript based on their feedback. Two overarching themes of the comments regarded nomenclature. We address these overarching themes generally below and then add clarifying responses to individual comments in the point-by-point response section. We have edited the abstract to emphasize the use of this analysis on application in atmospheric chemistry to avoid suggesting this is a novel technique.

Theme 1: Introduction of SKMD as a new formula/name

Both Dr. Fouquet and Anonymous Referee 2 expressed concerns regarding the novelty of SKMD (as compared to REKMD) and moreover expressed ideas about if analyses such as REKMD/SKMD should perhaps fall under a more description as "Kendrick analysis." We agree with the referees on this latter point and our original intention was to present the analysis as a generalization of REKMD not an entirely new analysis. However, since the results are not identical under all conditions, we chose a new name to avoid ambiguity.

To address this point, we have modified the manuscript to use the phrase "generalized Kendrick analysis" throughout except when referring to previous REKMD publications. This term is aligned with the suggestion made by Dr. Fouquet to refer to the idea as "Kendrick analysis" with the addition of the word "generalized" to avoid unnecessary confusion with the traditional Kendrick analysis that many within the mass spectrometry and atmospheric chemistry communities are familiar with. Eventually, it may be appropriate to drop the term "generalized," but we prefer to keep it at the present to provide clarity. We have adopted the formulation of "traditional" Kendrick mass transformation as

$$m_K\left(\frac{m}{z}, R\right) = \frac{m}{z} \times \frac{A(R)}{R} \quad (1)$$

Where we have incorporated new terminology m_K for the Kendrick mass rather than KM to better align with IUPAC standards for mass naming. We also replace the round(R) term with the equivalent, but more appropriate, nucleon number (A) of the base unit. By using nucleon number, m_K remains a mass but on a rescaled system. The modified Kendrick transformation then becomes:

$$m_K\left(\frac{m}{z}, R, X\right) = \frac{m}{z} \times \frac{X}{R} \quad (2)$$

where m_K is the modified Kendrick mass, X is an integer value. We present the REKM equation the same as before to be consistent with the past publications describing that technique. We believe that using Equation 2 as opposed to the formula introduced in REKM analysis is preferable because it relates to the "traditional" Kendrick analysis in a clearer way and allows to user to better understand the transformation occurring. We note that the main text explains that generalized Kendrick analysis is identical to REKMD for a subset of X , and any rational X in REKMD can be reproduced with an integer X in GKA.

We want to emphasize that our main objective in this manuscript is to describe the applications of this type of analysis to the subset of the mass spectrometry community focused on atmospheric chemistry/physics. As noted by Referee 2, such measurements are becoming increasingly common. Due to the increase in resolving power and time resolution, non-targeted analysis is becoming increasingly prevalent, and we believe that the community will benefit from considering new ways of visualizing the data that provide increased chemical information. Moreover, as the complexity of the mass spectra grow, so does the analysis time required. As such, our original goal was to focus on the advantages that this type of analysis provides in terms of visualization and assignment of chemical compositions with particular emphasis on quantities that are of potential interest to atmospheric chemists. We feel that it is appropriate to publish this analysis in this journal as a stand-alone contribution as the analysis method represents a new way of thinking about Kendrick mass transformations to many within the community, rather than

attempting to incorporate the explanation within a manuscript that also addresses questions of atmospheric chemistry. We recognize that the original structure of the manuscript may have obscured these goals and we have reorganized the discussion to clarify these points.

Theme 2: Use of mass defect terminology

Drs. Fouquet and Gonin both commented on how the term “mass defect” as used in this manuscript, and by the mass spectrometry community more generally, is inconsistent with the definition from nuclear physics. As pointed out by Dr. Gonin and identified by others (Pourshahian, 2017), the definition of “mass defect” used by the mass spectrometry community should be termed mass excess to be consistent with the definitions from nuclear physics. However, as pointed out by Dr. Fouquet, mass excess does not strictly apply when the chemical composition is unknown. As we highlight in the introduction, modified Kendrick analysis is suitable for use with unidentified ions and thus the term mass excess will also be incorrect in certain applications. We have added the following discussion to the manuscript to highlight these points.

Previous literature has referred to the difference between the nominal and exact mass as mass defect (Kendrick, 1963; Craig and Errock, 1959). However, we note that the terminology of “mass defect” in this application is incorrect as mass defect refers specifically to the difference in mass between the sum of the individual proton and neutrons in an atom and the actual mass of the nucleus due to the atom’s binding energy. The difference between a molecule’s integer mass and exact mass is due to how the mass scale of atoms is defined not solely due to the binding energy of the nuclei, therefore, “mass defect” should not be used (Pourshahian, 2017). For example, the mass defect of a ^{12}C atom in mass spectral analysis is 0 amu, while in physics it is 0.1 amu. Alternative names such as mass excess could be used in lieu of mass defect, though the previous adaptation of “mass defect” within the mass spectrometry community makes this transition difficult. Therefore, while we keep the term “mass defect” in this work, we have adopted the term generalized Kendrick analysis (GKA) when referring to quantities similar to those previously referred to as Kendrick mass defects. We do this to attempt to move away from incorrect terminology while also noting that with the use of the round function in Eqs. (2) and (3) the result is not technically a mass.

Point-by-point responses to Referee and Community Comments

We have copied the reviewer’s comments below in **bold**. Our responses are below each comment, in **blue**. Any additions to the text are *italicized* and removals are ~~struck through~~.

- Reviewer 1, Thierry Fouquet:

Apologies for being late with my comments, I had trouble submitting my post.

I read this article with great interest and did appreciate its form and the application of Kendrick analyses for a new type of mass spectral data and unusual instrument. I also acknowledge the efforts of the Authors to try finding differences between their proposed “scaled Kendrick mass defect” SKMD and the existing “REKMD” or “traditional” Kendrick analysis. If I strongly support new applications of this easy and powerful data processing/visualization tool and its implementation in more programs, I am nonetheless circumspect about the added value of the SKMD formula justifying the introduction of a new name and a seemingly new concept. Based on my comments below, I do strongly recommend not to try adding a new term to a long list of wrongly named methods but keep using either “REKMD” (also wrongly named, but published before) or better simply “Kendrick analysis”, and shift the focus of this draft from the

differentiation SKMD / REKMD which is non-existent to the application of the Kendrick analysis to their unusual data, providing examples of ion series assignments, separation of ion series of interest from congested mass spectra, deisotoping, binary comparisons,

1) The formula of “SKMD” look different from the formula(s) of “REKMD” proposed in the original article, but this notation $KM=m/z^*x/R$ has already been reported in *J Mass Spectrom.* 2019;54:933–947 (doi: /doi.org/10.1002/jms.4480), sadly not cited by the Authors.

We thank Dr. Fouquet for bringing this publication to our attention. We have included this in our references and added the following text to the manuscript after we introduce the form $KM=m/z^*x/R$:

Note that this form of the equation has been demonstrated before in polymer mass spectrometry (Fouquet, 2019). However, its applications and advantages with respect to visualization and ion assignment as used in atmospheric chemistry has yet to be identified and discussed.

2) The case $x=1$ in the formula $KM=m/z^*x/R$ has already been explained in the same article, and named “remainders” by several Authors (e.g. *Anal Chem.* 2019;91(10):6479–6486 or *Anal Chem.* 2018;90(14):8716–8718). The only change introduced by this preprint would then be the range of values $1 < x < 2/3R$, and $x > 2R$. Looking at the plot $\Delta SKMD$ and $\Delta REKMD$ in the Supporting Information (Fig S1), it seems that the expansions provided by SKMD are already achieved by REKMD in its linearity range. I have not been convinced by the examples of $x=4$ and $x=40$ (not truly REKMD in terms of range of divisors) in terms of gain of visualization / separation as compared to the other plots ($x=17, 20, 24$) which are truly REKMD. Can the Authors find a case where values of $x < 2/3R$ or $> 2R$ provide a truly unique separation capability not achievable by REKMD with integer or non-integer x ? I haven't found any case myself yet.

We appreciate the reviewer for bringing these publications describing the situation when $X=1$. We have included them into our manuscript and added the following text:

The increased expansion or contraction ability of GKA compared to REKMD may not be useful for every set of mass spectrometric data, as the range of X available in REKMD may be sufficient. Though as will be discussed, the increased expansion leads to an increase in separation of points and easier visualization of the different ions.

We agree that the expansions of REKMD are linear within the values of $2/3R < X < 2R$. We provide an alternative equation for the Kendrick analysis for situations in which larger or smaller values of X would be helpful. We respectfully disagree that $X=40$ in Fig 4C does not provide an improved visualization. The increased expansion does lead to more separated points that improves visualization for the entire mass defect range.

3) The formula $KM=m/z^*x/R$ is still fundamentally a traditional Kendrick change of basis as proposed by Kendrick a while ago, simply choosing x instead of $\text{round}(R)$ as the new reference mass. The formula $KM=m/z^*\text{round}(R)/R$ is a basic “rule of three”, setting the mass of R at an integer value $\text{round}(R)$ to define a new reference instead of the IUPAC convention $m(12C)=12$, recalculating other mass accordingly. In the SKM formula, the Authors choose to set the mass of 160 (or other moieties) at 2, or 6, or 40 or any integer instead of 16 to define a new mass scale. The concept does not vary from what Kendrick proposed, so does this really deserve a new name? I do agree that the same question should have been raised when the concept of REKMD has been introduced. That is the reason why I am strongly in favor of calling the whole method a “Kendrick analysis” with no other mention.

m/z 160 IUPAC = 15.9949 --> m/z 160 Kendrick base = 16

m/z IUPAC --> rule of three --> m/z Kendrick base = m/z IUPAC *16/15.9949

m/z 160 IUPAC = 15.9949 --> m/z 160 Kendrick base = x

m/z IUPAC --> rule of three --> m/z Kendrick base = m/z IUPAC *x/15.9949

We believe our response under “theme 1” at the beginning of the document addresses this comment.

4) REKMD and its latest variations as reported in detail in the same article J Mass Spectrom. 2019;54:933–947 takes the charge state of the ions into account (simply adding an integer to the formula to cancel z) so it is applicable for multiply charged ions, while SKMD in its current form would deal with singly charged ions only. More importantly, this additional integer in the formula of KM also made possible the generation of an infinite number of Kendrick plots with a pseudo continuous coverage of expansions using non-integer x (ie nearly 0 step between expansions allowing the finest tune to separate series) while SKMD does only provide fixed expansions varying linearly with a step dictated by the value of R itself, and no finer control available (the larger R, the larger the step).

This work focused on mass spectrometry data that only has singly charged ions, therefore we did not add an additional multiplicand to account for the charge state. If necessary, the analysis can easily be modified to incorporate multiply charged ions (Fouquet, 2019), but since most applications within the atmospheric community involve singly charged ions we do not consider multiply charged ions here. If this adjustment is made, the mass of the electrons will also need to be incorporated in the calculation. Additionally, we have included the following text to address the second point in the comment:

As previously mentioned, REKMD can use select rational values of x. To maintain horizontal alignment of homologous series, only rational values of x satisfying $x \cdot \text{round}(R/x) = \text{integer}$ are allowed (Fouquet, 2019). Substituting that condition into Eq. (3) results in

$$\text{REKMD} \left(\frac{m}{z}, R, x \right) = \frac{m}{z} \times \frac{\text{integer}}{R} - \text{round} \left(\frac{m}{z} \times \frac{\text{integer}}{R} \right) \quad (6)$$

Showing that all expansions achievable with REKMD using rational values of x can also be achieved with GKA. With GKA, pseudo-continuous expansion becomes possible without introducing extra multiplication factors as is necessary in REKMD (Fouquet, 2019) and thus it is appropriate to consider GKA as a generalization of traditional Kendrick analysis and REKMD.

5) REKMD or its variations has already been implemented - but not called as such - in numerous programs, free or commercial, such as MZMine2, MSRepeatFinder, SpectraScope, Kendo, or Mass Mountaineer not to mention in simple Excel spreadsheets. These programs do not mention any REKMD but simply incorporate the divisor “x” which can be changed by the user. It is a wonderful idea to keep implementing the Kendrick analysis with all its variations in other programs, but would that require to use a new name and a seemingly new concept which in fact produces the same results as those already reported ? Would it be clearer for users to have a tool called “Kendrick analysis” with no S or RE or no mass defect (cf last comment below), but simply this “x” textbox to play with the change of basis and expansions ?

We thank the reviewer for bringing these other programs to our attention. We agree that using different names should be avoided as discussed in “theme 1” at the beginning of the document. We have also added X to the panel to clarify that input is the scaling factor.

6) As a last reason not to introduce another (S)KMD term, and as pointed out by one reader of the community in his comment, several Authors strongly recommend not to use the “KMD” term anymore. The values we are dealing with are not mass defects – and not mass excess either – but fractional mass calculated with no a priori knowledge about the ions. This point has been greatly explained by the Authors in this preprint, as opposite to other plots such as Van Krevelen diagrams which require the elemental composition of ions to be known prior to their generation. The calculation of true mass defect/excess ALSO requires the elemental composition of the ions to be known. The y-values of Kendrick plots are computed a) without knowing the compositions of the ions but only their m/z , and b) are comprised between -0.5 and 0.5 (or -1 and 0, or 0 and 1) regardless of m due to the aliasing of the formula $m - \text{round}(m)$ (or floor or ceiling) while the mass defect/excess would keep increasing/decreasing with the number of atoms in a molecule above or below (-)0.5

Please see our response under “theme 1” at the beginning of this document.

Anonymous Reviewer 2:

Overall, I think that this manuscript provides a valuable demonstration of a “Kendrick analysis” (using terminology from RC1) that can be applied to complex mass spectrometry data (in this case a VOCUS-PTRTOFMS) without knowledge of the ion elemental composition to improve data visualization towards assisting in ion assignment, revealing chemical homologues, and potential chemical trends. The manuscript is generally well-written and straightforward, thoughtful about introducing KMD, REKMD, and SKMD, and definitely proves through the figures the potential for using effective scaling for allowing greater insight into atmospheric chemical measurements (e.g. the separation of nominally odd/even IUPAC mass ions into odd nitrogen containing or no/even nitrogen containing formulas). That being said, I agree with RC1’s concerns that the main formulation of SKM/SKMD not truly being novel per prior publication in Fouquet (2019) necessitates reframing how the manuscript is presented/worded. After this, I still see the manuscript being publishable and of interest to the atmospheric measurement community for bringing attention to a potentially valuable data processing method for the torrent of mass spectrometric data being collected in recent times.

We thank the reviewer for their thoughtful comments. As discussed in “theme 1,” we have reframed the manuscript to not focus on the novelty of this analysis method as we agree it has been previously published in other mass spectrometry communities. We have renamed the equation to generalized Kendrick analysis to avoid the misconception that this is a novel analysis method. We also have included more references and discussion of the references that we were missing from other mass spectrometry communities to further illustrate that the purpose of this paper is to introduce this method into the atmospheric sciences community where, to our knowledge, it was not previously known.

Specific Comments:

Line 172: the introduced term “reduced fraction” representing $X/RIUPAC$ is not the most intuitive given values of X (as exemplified in the cases of $X=20, 24, 40$) with RIUPAC as ^{16}O results in expanding the mass scale instead of contraction—this seems more like a scaling factor whether

enlarged or reduced. Consider changing this term and revising throughout the rest of the discussion.

The term “reduced fraction” was not meant to represent X/R per say, but rather the numerical value of the quantity expressed as a reduced fraction (i.e., reduced fraction of 20/16 is 5/4). We agree that the use of the term reduced could be confusing as we are not talking about the reduction of a mass scale, but rather a fraction. As reduced fraction is the proper mathematical term for the concept we are discussing, we are hesitant to change the wording. Therefore, we have edited the line be clearer as follows:

Figs. 1a and 1b compare $KM(^{16}O)$ and $GKA(^{16}O, 24)$. For the $GKA(^{16}O, 24)$ analysis, the mass scale is expanded by a factor of $\sim 3/2$ (from the approximate reduced fractional value of X/R, or $24/15.995$).

Additionally, we have added the following text:

As such, the number of groupings from a certain transformation is the reciprocal of the denominator. Note that the reduced fractional value of X/R does not determine the amount of contraction or expansion of the mass differences, but rather determines the number of individual groupings of related ions.

Line 303: The text would be enhanced to include more information such as the caption for Figure 7. That is, explain here that these points in Fig. 7a are not just simply omitted in Fig. 7b, but they would not appear in Fig. 7b because they would not fall within the SKMD range after m/z transformation.

The following text has been added:

The grey points in Fig. 7b are points that appear within the mass defect space using a traditional KMD analysis but are not visible with GKA (Fig. 7a) since they are shifted to another area of the mass defect space.

Lines 241, Line 251, Line 296: By using “REKMD” in these section headings, it signals to the reader that co-authors intend SKMD to be a sub-method/type of REKMD, yet the language earlier in text introducing SKMD as a concept makes it seem that it should be distinct as it is used in the analyses in these sections. Thus, I would have expected these section titles to be “SKMD” instead of “REKMD”. Given the comments from RC1 on nomenclature/reframing the paper generally as a “Kendrick analysis,” just be consistent with the chosen framing in the revised version.

We have removed REKMD from all section headings and referred to the technique as generalized Kendrick analysis to avoid confusion.

General comment on Figures 3, 4, 5: For further connection with text and enforcing of how mass defect analysis allows for visualization of homologous series, it would be helpful for labels on the figures pointing out the chemical families and their generic chemical formulations if possible.

We have tried to include this additional information in the figures but found that the additional text makes the figures very difficult to read and easily understood, which is why we elected to color points in certain figures by some property relating to the chemical formulas. Coloring by user-set parameters is achievable with the GUI we present. In addition, we have included a modified version of Fig. 3 in the SI (Fig. S4) that shows two zoomed in sections of each panel to show the chemical formulas related by ^{16}O along horizontal lines.

Technical Corrections:

It would be helpful perhaps in Fig. S1 to include vertical lines associated with the bounds of X presented in Eq. (7), which brings more focus to the linearity and equivalency of the two methods within those bounds.

These lines (and horizontal lines) have been added with accompanying text in the caption.

Figure 2 captions for c and d have inconsistent X values with those in subpanels of figure. Caption text for c) should have X = 24, d) should have X = 20.

Corrected

Line 148: Extra space before period should be deleted.

Corrected

Line 180: Change “result” to “results”.

Corrected

Line 278: Add period after “respectively.”

Corrected

Figure S4: Add panel labels to figure

Corrected

Public Comment 1, Marc Gonin:

There is an error in nomenclature in this paper. This concerns the quantity "mass defect" that is used in a way which is not in line with the usual scientific definition.

mass defect:

- is the difference between the masses of the free elementary particles and the bound elementary particles of a nucleus, and atom or a molecule.
- it is always positive and never negative
- it represents the binding energy by $E = mc^2$
- see https://en.wikipedia.org/wiki/Nuclear_binding_energy

mass excess:

- is the difference between the mass and the nucleon number x Da (= nominal mass) of nuclei, atoms or molecules
- it is positive for atoms with low binding energies per nucleon, and negative for atoms with high binding energies per nucleon.
- https://en.wikipedia.org/wiki/Mass_excess
- conceptually, if a defect is large the mass must be small, the excess must be small

The quantity you referring to is therefore mass excess, not mass defect.

Mass spectrometrists have a long history in getting this wrong. It is now even wrong in textbooks and reference materials. However, in my opinion this should not serve as an excuse to accept this faulty nomenclature. Mass defect and mass excess are caused by nuclear physics and therefore mass spectrometrists should accept and use their nomenclature.

It is not helpful if every branch of science develops their own jargon. This hinders mutual understanding and promotes the view of science as isolated ivory towers. Science is under large criticism today, and we should not promote criticism by using sloppy nomenclature. Scientific rigor is the only answer to this criticism

We thank Dr. Gonin for his insightful comments on this manuscript. We agree that we should move away from incorrect terminology in this and future publications. However, “mass defect” is engrained in the mass spectrometric community and we do not want to cause more confusion by introducing new terms to this manuscript. We are also hesitant to use the terminology of “mass excess” as described in our “theme 2” response. Therefore, we avoid the term “mass defect” by referring to the values as results of GKA.

Public Comment 2, Marc Gonin:

When discussing mass excess or mass defect, the quantity mass should be used in the plots, not the quantity mass/charge. This in spite of mass spectrometers measuring mass/charge. The relevant quantity in these discussions is mass.

We elect to keep the quantity mass/charge instead of mass to be consistent with units familiar in the mass spectrometry community. Additionally, in the case of unidentified ions, the value measured by the mass spectrometer, and what is used in generalized Kendrick analysis is not mass but rather mass/charge. As this analysis can be used on unidentified ion, we believe mass/charge is more appropriate for visualization purposes.

References

Fouquet, T. N. J.: The Kendrick analysis for polymer mass spectrometry, *J Mass Spectrom*, 54, 933–947, <https://doi.org/10.1002/jms.4480>, 2019.

Pourshahian, S.: Mass Defect from Nuclear Physics to Mass Spectral Analysis, *J. Am. Soc. Mass Spectrom.*, 28, 1836–1843, <https://doi.org/10.1007/s13361-017-1741-9>, 2017.