

Comments in bold font, replies in regular font. Sentences added to the updated manuscript are written in italics.

RC1, General comment

This study presents a technical advancement toward automating peak fitting and formula assignment in mass spectrometric analysis, a task traditionally requiring significant manual effort. The authors developed an algorithm intended to streamline this process by integrating weighted least squares fitting, a modified Bayesian information criterion, and iterative formula assignment. Their approach aims to deliver a preliminary list of likely peaks that can later be refined as needed, thus reducing the time analysts typically spend on labor-intensive, manual identification. The algorithm was tested on gas-phase CIMS and aerosol AMS datasets, and showed comparable accuracy to manual fitting in many cases. However, as with manual methods, lower-intensity signals and interferences from adjacent peaks presented challenges, which led to occasional erroneous assignments. The study's main output — the algorithm — is reported to be undergoing integration into the Tofware analysis software, making it accessible to a broader user base and thus potentially transforming routine data processing workflows. Overall, the work represents a valuable contribution to the field, promising to free up researchers' time for data interpretation rather than time-consuming data processing tasks. I recommend its publication in AMT, considering the following comments are adequately addressed.

We thank the reviewer for their insights and positive response. The specific comments are addressed below.

Specific comments

Q1. I assume the peak shape function follows a Gaussian distribution? Since position is defined based on the peak shape function, variations in shape could impact the algorithm's reliability.

It is correct that this definition may impact where the actual fit is made, and it is important that the peak shape used, and its position, are consistent. The peak shape is a custom function that is extracted from the dataset in established software packages by selecting isolated peaks, normalizing them, and averaging their shapes. The position of the peak is commonly defined as the maximum of the peak shape function. However, as long as the same peak shape is used for the algorithm as was used for the mass calibration, this

should not impact the reliability of the results, since the mass axis will be defined in accordance with the defined position relative to the shape.

We added the following two sentences to clarify this in section 2.1.1: *The position of the peak is commonly defined as the maximum of the fitted peak shape function. It is important that the definition is consistent between mass calibration and the fitted ions to accurately determine the mass-to-charge ratio of the detected ions.*

Q2. More detail on the rationale for default parameter values, especially for critical values like n_{max} and the parameter A, would be helpful. How robust are these defaults across different instruments and sample types? It would be useful to know how sensitive the fitting results are to these parameter settings and if there is a straightforward way to optimize or validate these parameters for different datasets.

We agree that straight forward parameters are important for practical use of an algorithm such as this. A user should not need to spend more time on figuring out input parameters than what is saved by adoption of the tool. n_{max} is simply the maximum number of peaks to attempt fitting. We believe someone working on mass spectrometry data will be aware of roughly what number would be suitable here, but the only cost of using a higher number would be longer computation time, so choosing a value a little too high should not be too much of an issue. We expect that the results presented in the manuscript would be quite similar when choosing any value greater than or equal to 10, given that the algorithm never fits more peaks than this. In this work the value 12 is chosen because it would be extremely optimistic to expect anyone to successfully identify 12 peaks at one mass in a spectrum like this. Even a value of 8 should not have a very large impact on the results, since very few masses get assigned that many peaks.

We now clarify the impact of a higher n_{max} value in section 2.1: *Picking a higher number than necessary only costs additional computational resources.*

The parameter A is automatically determined as described on lines 115-121 and in appendix A3. This determination is designed to be as independent of instrument type and dataset as possible. In fact, the very reason A is determined in this manner is to make the algorithm as general as possible. There may still be cases where the method is not perfect, and the algorithm may struggle with certain datasets because of this, since it is difficult to establish a general method without a large variety of datasets. If there are cases where the method is not working in a general manner, it does need to be improved over time. With more users providing feedback improving the generality should be easier.

The same is true for other default values. As a rule we don't believe the default values to be detrimental to algorithm performance in any case, but

some instances may come up where they do require some tweaking. This also is something that requires ample feedback from users, and diverse use cases to identify.

We added clarification on the purpose of the fitting procedure to determine A in section 2.1.2: *The optimal value of this parameter may vary between datasets, so to make the algorithm as general as possible it is automatically determined before the peak identification portion of the algorithm.*

Q3. The current formula list appears to be derived from existing datasets validated by specific instruments, which are selectively sensitive to certain groups of compounds. While these formulas are relevant to particular compounds, they don't encompass all possible combinations of elements that adhere to established Chemical bonding rules. Given the complexity of organic carbon mixtures in the atmosphere, expanding the list to include additional elements beyond C, H, O, N, and S, and more importantly, to include all possible formula combinations that obey the valency rules, will be a crucial step for future development.

Yes, we believe that the potential formula list is very important for practical use of the algorithm. A general list may be useful for newer use cases, but for commonly used mass spectrometers more restrictive lists (such as the ones presented in the manuscript) would likely be more useful. With increased use of this algorithm we hope to provide some type of library where users can find, or even submit, lists for different tools and applications. With the algorithm being able to label peaks "unknown" when there is no suitable compounds, the cost of using a list which does not include all of the detected compound is not so high, and a more restrictive list could avoid having too many formulas with very similar masses.

Choosing the sparsity of the list may also be a useful depending on the goals of the analysis. Perhaps a user is only interested in some class of peaks, and therefore chooses to omit other classes of compounds. For example, when analyzing the gas phase dataset presented in the manuscript it may not be necessary to include fluorinated compounds at all, since they have a very different mass defect from the pure C, H, O, and N containing compounds targeted in the original study.

Q4. Page 13, line 301: Following the previous comment, implementing the odd nitrogen rule would serve as a valuable criterion for automatically excluding incorrect formulas as discussed here.

We agree that the odd nitrogen rule may be useful for many datasets. This is the reason why "Formulas with exactly 3 Nitrogen and an odd number of Hydrogen" are not included in the gas phase list (note that one nitrogen is expected to come from the adduct ion). However, the nitrogen rule does not apply for radicals, and we do actually expect to see some of those in the gas

phase dataset. This is why the nitrogen rule is only enforced for the "rarer" 3 nitrogen formulas.

We now clarified our reasoning in Appendix A2: *The last criterion on the list is to make sure the formulas with three nitrogen follow the nitrogen rule. The nitrogen rule is not considered for other numbers of nitrogen atoms since formulas with 0–2 nitrogen are more abundant, and even radicals can be observed.*

In the particle phase data we detect mostly fragments of molecules that have been vaporized. Therefore we cannot restrict the data very much based on chemistry for that dataset.

We also added clarification on the decisions for generating the particle phase dataset: *Both the constraints and the parts used to build the formulas in this list are less general than for the gas phase dataset. This is because the AMS mostly detects fragments of molecules so valency rules cannot be used to constrain the list.*

Q5. Given that low-intensity peaks are prone to interference and misassignment, further explanation on strategies for handling such peaks would strengthen the method. A more rigorous treatment of noise and background signals, including methods for background subtraction, and options to customize baseline input, could improve accuracy.

The second referee asked a similar question and this reply aims to answer both of these questions. The same reply is given in response to Referee 2 Question 16.

As mentioned in the comment, background is important, but isolated low-intensity peaks are not that challenging. The real problems are the low intensity peaks overlapping with other (potentially also low intensity) peaks. Here the main problem in our experience is the compounding errors from mass calibration and peak shape from the other nearby peaks. These errors will add up from all the nearby peaks, and additionally a small error in a big peak may have a huge impact on a small peak. In addition, there is the noise from counting statistics contributed by the overlapping peaks, further decreasing the signal of the low-intensity peak relative to the noise.

For most of these issues the solution is to better define the parameters i.e. minimize calibration errors, and errors in peak shape. Also removing as many unrealistic compounds as possible from the potential list may be helpful, but this runs the risk of missing something of scientific importance. Improving the accuracy of these parameters is not straight forward. Some of it relies on the expertise of the analyst, and their familiarity with the dataset, and some on the available tools. With user friendly tools, even an inexperienced analyst could achieve good results. We have not discussed these parameters extensively in the manuscript since they are treated as input parameters for

the algorithm. However, the two don't have to be treated separately forever. Although outside the scope of this work, including some refining of these parameters in the algorithm workflow could lead to significant improvements in results. This would be the next step in improving this algorithm further.

It is also important to note that this is not an algorithm specific problem. There is a limit of what can be inferred from the provided dataset due to the signal-to-noise ratio, and unavoidable uncertainty in parameters, such as mass calibration. Regardless of if the fitting is done manually or automatically, there will be some signals that cannot be assigned with confidence, or cannot be distinguished from the noise at all.

Concerning the baseline in particular, the algorithm seems to handle it quite well. A baseline is fit at each mass, as a constant background. Fitting the baseline along with the peaks is important, because the algorithm will otherwise be highly sensitive to the input baseline, and small errors will result in very big issues with the low- or even medium-intensity signals. However, baselines are not always constant, so the option to fit a linear, or even quadratic baseline could improve this approach further. Currently, non-constant baselines can be handled by the user inputting a baseline, and the algorithm adding a fitted baseline on top of that.

The following was added to the Discussion (section 3.4): *As evidenced by the results, the most challenging problem with peak identification are the peaks with lower signal, overlapped by other peaks. These peaks are more significantly impacted by errors in mass calibration, peak shape, or assignments since the errors of overlapping peaks add up. The difference in signal also means that a small inaccuracy in the peak shape for a large peak can lead to a large impact on a smaller peak in the vicinity. Therefore we believe the primary means of improving the results in general, but most significantly for the lower intensity signals, is to improve the methods to accurately determine these fitting parameters.*

We also added the following paragraph a bit later on in the same section: *However, since peak fitting is a statistical tool, there will always be an inherent level of uncertainty. Whether the identification is done manually or by an algorithm, there will be some peaks that cannot be identified with the desired confidence. Where this limit is encountered depends on the input data. The goal of the algorithm is to save time, so we do not think it is necessary to demand it to be able to identify more peaks than manual fitting.*

Q6. I am not sure if isotopic checks are currently incorporated in the algorithm, but incorporating an optional check for isotopes could reduce misassignments, particularly for elements with non-standard isotopic distributions

We did consider this as a potential tool, and we did some tests for this. The topic of isotopic checks is discussed like this in the manuscript:

(Section 2.1): "The algorithm does not currently check if the isotopes for an assigned formula are present. This may be a useful future improvement, but testing showed it would very rarely be useful in the datasets tested here."

To clarify, we have only found one single case in the results where an isotope check would have changed a peak assignment.

(Section 3.4): "Another future improvement would be for the algorithm to reconsider formulas, whose isotopic signals do not match the data. As mentioned previously, this was found to be relevant very rarely during testing. In part due to most organics having fairly similar isotopic patterns and in part because the algorithm mostly misidentified peaks with comparatively low signals. Even for datasets containing halogens or other elements with isotopic patterns that deviate from organics, the different mass defect should result in accurate identification of these formulas in a majority of cases. However, this may be an improvement for future consideration."

We hope this adequately addresses the concerns of the referee.