The reviews of our manuscript are thorough and well-considered. We would like to thank the reviewers for their careful reading and valuable comments, as well as for the improvements they help us make.

All the suggestions and comments from Referee 1 are addressed below point by point in bold text, followed by our responses in non-bold text. The corresponding revisions to the manuscript are marked in blue. All updates to the original submission are tracked in the revised manuscript.

Wu and colleagues present the novel application of a data treatment technique (fuzzy *c*-means clustering) that reduces the complexity of a data set for the interpretation of atmospheric mass spectra. I consider the exploration of novel techniques for the extraction of information from increasingly more complex and information rich mass spectra to be of interest to the atmospheric science community and the readership of AMT. The manuscript strikes a balance between being an introduction to the technique and the demonstration of the technique on an established chemical system. I consider the chosen format and presentation as appropriate for the presentation of a new approach, and can see how this manuscript could serve as a reference for future studies using the technique.

I enjoyed reading this manuscript and would welcome a publication after some minor points have been addressed.

Thanks for the positive comments. It’s appreciated.

Could you touch briefly on the practical implementation of the algorithm? I do not readily find what software you used. What is the code availability? What are typical run times of the algorithm?

Response: The fuzzy c-means clustering was implemented using the open-source scikit-fuzzy (v 0.4.2) package (https://pypi.org/project/scikit-fuzzy/) in Python. For each case, the algorithm was repeated 50 times to get a robust result, and each run took several to tens of seconds depending on the number of clusters.

In the revised manuscript in L293-294, a sentence "In this study, the FCM clustering was implemented using the open-source scikit-fuzzy (v 0.4.2) package (https://pypi.org/project/scikit-fuzzy/) in Python." has been added to address this point.

I did not check the correctness of the equations, but encourage the authors to double (or even triple) check the manuscript for typos in the equations before the final publications.

Response: Thanks for this comment. We have double checked all the equations in both the main text and the supplement. Typos have been revised.

Else, I found few typographical errors, only, which will be taken care of in copy-editing, and limit my comments to content observations only.

Response: Thanks for this.

Paragraph 3.1.1: unclear on what data set you "ran the FCM algorithm 50 times", please clarify.

Response: We apologize for the unclear expression. In this study, except for results shown in Sect. 3.2.2, all others are results from the FCM analysis of chamber data. In Sect. 3.1.1, the FCM algorithm was applied to chamber data to find out the optimal number of clusters.

In order to make this point clearer, the original sentence has been revised to "To explore the effect of cluster number on partition results, ~~we ran the FCM algorithm 50 times for each c in the search range and calculated the corresponding CVIs~~ we applied the FCM algorithm to the chamber data with c varying from 2 to 10. For each c in this range, the algorithm was run 50 times and the selected CVIs were calculated accordingly for each repetition.".

Paragraph 3.1.3: expand on the accuracy of *m**. Is 1.42 and 1.52 significantly different or essentially the same? Generally, what can be considered as different?

Response: That is a good point! The value of *m* defines the fuzziness degree of a clustering, which affects the convergency of the algorithm and the separation of clusters. Generally, the larger *m* is, the fuzzier the FCM clustering results would be. We have checked the clustering outcomes with *m* = 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0. The patterns of cluster centers are almost identical with different *m*. However, for different *m*, some of the membership degrees of objects to cluster centers changed to some degree (but less than 50% at most), which implies that *m* does affect the clustering outcomes. In this regard, we would say selecting *m* = 1.4 or *m* = 1.5 makes a difference.

In addition, the major purpose of this section is to provide a method to determine the "optimal" fuzzifier value for a given data set, rather than using the default value of *m* = 2, which has been proven to be inappropriate in many conditions (Huang et al., 2012; Hwang and Rhee, 2007; Schwämmle and Jensen, 2010; Yu et al., 2004; Zhou et al., 2014). In this study, we intend to provide a method to determine the optimal value of *m*. The exact value of *m** is not what we really care about.

Fig. 4: I was a little thrown off by the missing x-axis label on the mass profile panels. Can you add a label, and/or label the dominant species in the individual panels? Also, consider changing the time axis label to elapsed time since start of experiment.

Response: Accepted. Due to limited space, species No. is added in the x-axis of panel (b) in Fig. 4, instead of chemical formulas. Specific chemical formulas of each species are listed in Fig. S7. The updated plot is shown below.

In addition, Fig. S4 has been revised accordingly. The time axis of Fig. 6 has also been changed to time elapsed since the start of experiment.

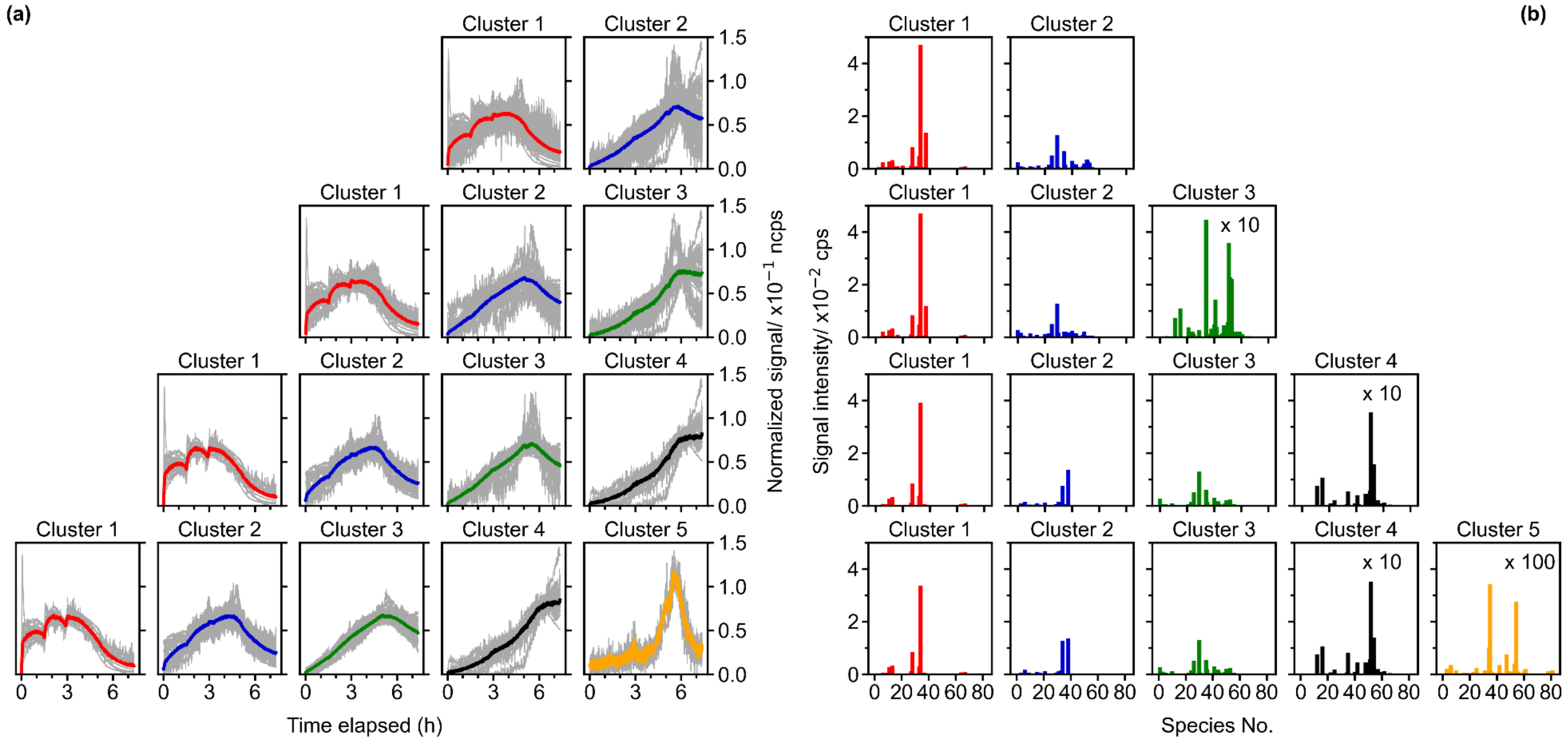


Figure 4. Results of fuzzy c-means clustering for chamber data with cluster numbers between 2 and 5: Time series (a) and mass profiles (b) of clusters for each solution (in row). The time series of cluster centers are shown as thick, colored solid lines, and the time series of species with the membership degree larger than 0.5 to the cluster are illustrated as thin, gray lines. The species number in panel (b) corresponds to species listed in Fig. S7 (in order of molecular mass).

L184: Consider rewording "and only formulas within an accuracy tolerance of 10 ppm and with reasonable chemical meanings were considered." to "and only plausible formulas with relative m/z deviations smaller than 10 ppm were considered"

Response: Done.

L290: Incomplete sentence. Missing "be"?

Response: Done. To make it complete and concise, the original sentence has been revised to "…, ~~what we believe to~~ the expected optimal number of clusters is determined.".

L300: Consider rewording "the right choice" to "may not always be appropriate".

Response: Done.

L428: "mathematically unsolvable". Probably better to say that there is no simple analytical solution.

Response: Done. The original sentence has been reworded to "…, and there is no simple analytical solution for the differential equations that describe Eq. 11 ~~are mathematically unsolvable~~.".

L467: Punishing function? While punishing seems to be used in some literature, it should maybe rather be penalty instead?

Response: Totally agreed on. Penalty function should be the right word. The "punishing function" in this sentence has been replaced by "penalty function". In addition, the improper expression in S1 has also been revised.

L493: reword “looks reasonable”

Response: Done. The original sentence has been changed as follows: “~~However, it looks reasonable~~ It seems more sensible to set the number of clusters to 5, as this is where *FS* reaches its local maximum and V_{SWCV} is significantly reduced and has the maximum curvature ~~which corresponds to the local maximum in terms of *FS*~~.”.

L512: Drop “As a quick reminder”

Response: Done. “As a quick reminder” in the original sentence has been replaced by “As mentioned before”.

L699: “mathematically”? Consider making the paragraph (especially lines 699-701) more concise.

Response: Done.

The original paragraph has been revised to “In this section, we utilize ~~will analyze~~ the five-cluster solution, identified as the optimal cluster number for our dataset (Sect. 2.3), to illustrate how to extract ~~to exemplify the functionality of FCM for extracting the~~ chemical and kinetic information from ~~underlying in~~ the mass spectrometric data based on the FCM analysis. ~~The five-cluster solution is chosen because $c = 5$ is the mathematically optimal cluster number determined for our dataset in sect. 2.3.~~ This does not necessarily mean that the five-cluster solution ~~we claim it~~ is superior over ~~to~~ other solutions~~, e.g., the six-cluster solution~~. However, as demonstrated ~~Besides, we confirmed~~ in ~~the~~ previous sections, ~~that~~ the FCM results exhibit consistent ~~general~~ features regardless of the ~~predefined~~ number of clusters predefined~~,~~ . Therefore, ~~so that~~ findings derived from ~~based on the analysis of~~ the five-cluster solution could potentially apply to ~~can hopefully also be generalized for~~ other cases.”.

L736: marker size: I appreciate the attempt to be specific about what the marker size represents. Please be fully specific by referring to the marker area or the diameter. This comment applies to a couple more figures in the main text and SI.

Response: The marker area of clusters is proportional to the sum of average signal intensity of all species in the cluster weighted by their membership degrees, and that of species is proportional to the average intensity of species over the whole experiment.

To make this clear, the original caption of Fig. 7 has therefore been rewritten to “Chemical properties of clusters from the five-cluster solution. The subplots show mass profile of each cluster (a), van Krevelen plot (b), and average carbon oxidation state of clusters (c), respectively. Different clusters are distinguished by color, and the color scheme follows that in Fig. 4. The marker area of clusters is proportional to the sum of average signal intensity of all species in the cluster weighted by their membership degrees. The species number in panel (a) corresponds to species listed in Fig. S7 (in order of molecular mass). Grey ~~circles~~ hexagons in panel (b) and panel (c) denote species identified by Br^- CIMS~~.~~ , and ~~The~~ the marker ~~size~~ area is proportional to ~~the square root of~~ the average intensity of ~~clusters/~~ species over the whole experiment.”

The captions of Fig. 5, Fig. 8 - 11, and Fig. S5 - S6 have been revised accordingly.

Fig. 11: Units for k

Response: k is a second-order rate constant in unit of cm^3 $molecule^{-1}$ s^{-1}. Figure 11 has been updated to include the unit of k. The updated plot is shown below:

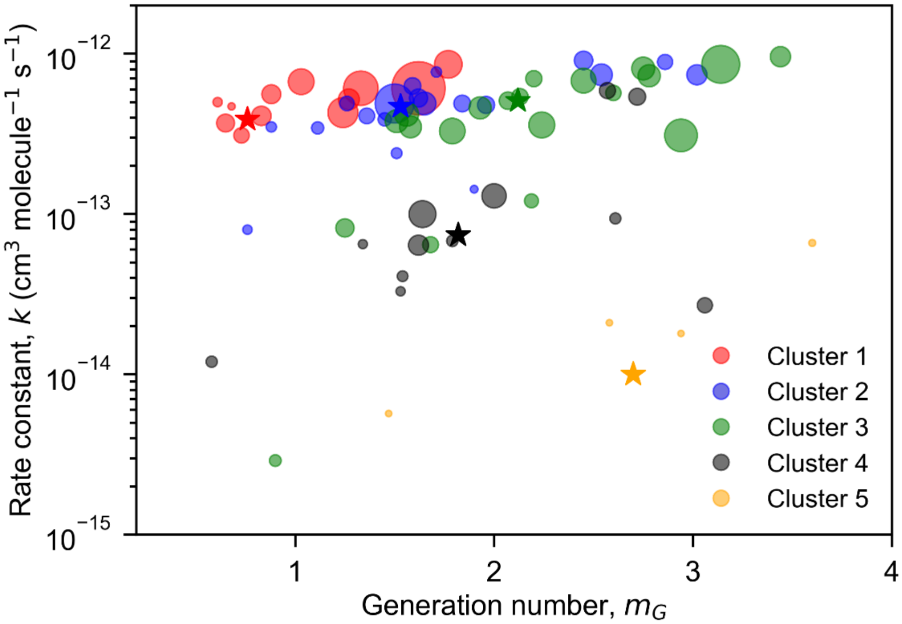


Figure 11. Fitted effective rate constant (k) and generation number (m_G) of the high-affiliation species of each FCM cluster. The cluster centers and members are denoted by color-coded

circles and pentagrams, respectively. The circle area is proportional to the average signal intensity of species over the whole experiment.

Supplementary information

Page 2: proposed >the< Kwon index.

Response: Done.

Page2: punishing function, same as main text.

Response: Done. “punishing function” has been replaced by “penalty function”.

Page 4: it’s --> it is

Response: Done.

Fig S1: axis labels cut off

Response: The size of this plot has been adjusted to ensure that it can be displayed completely. The updated version is shown below:

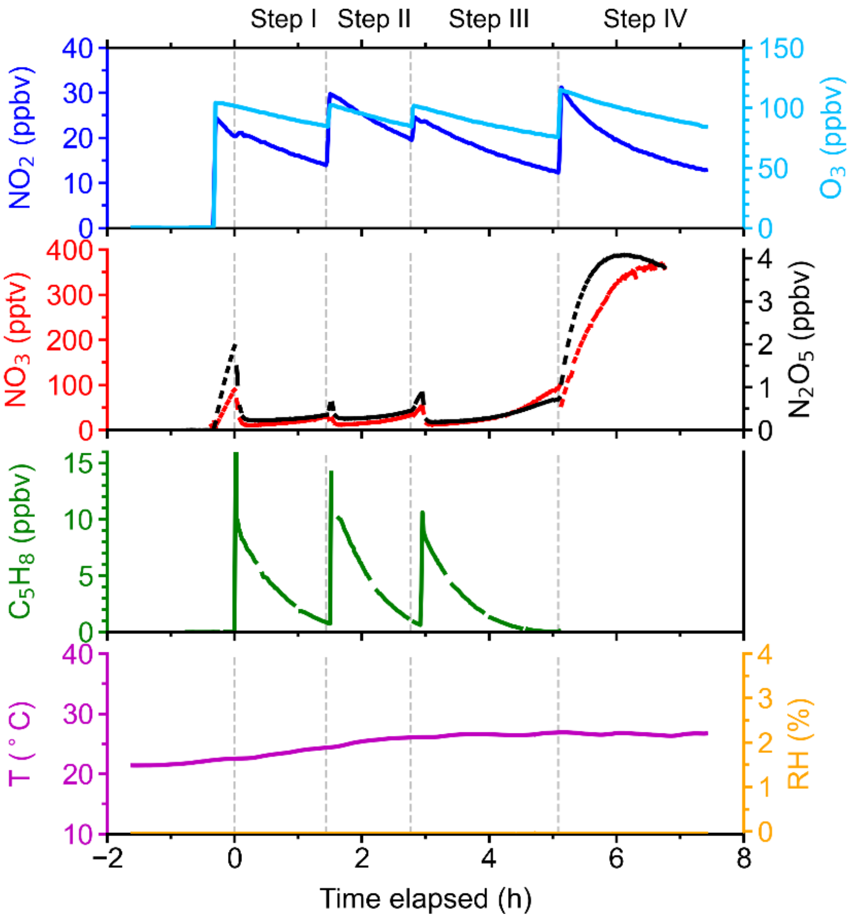


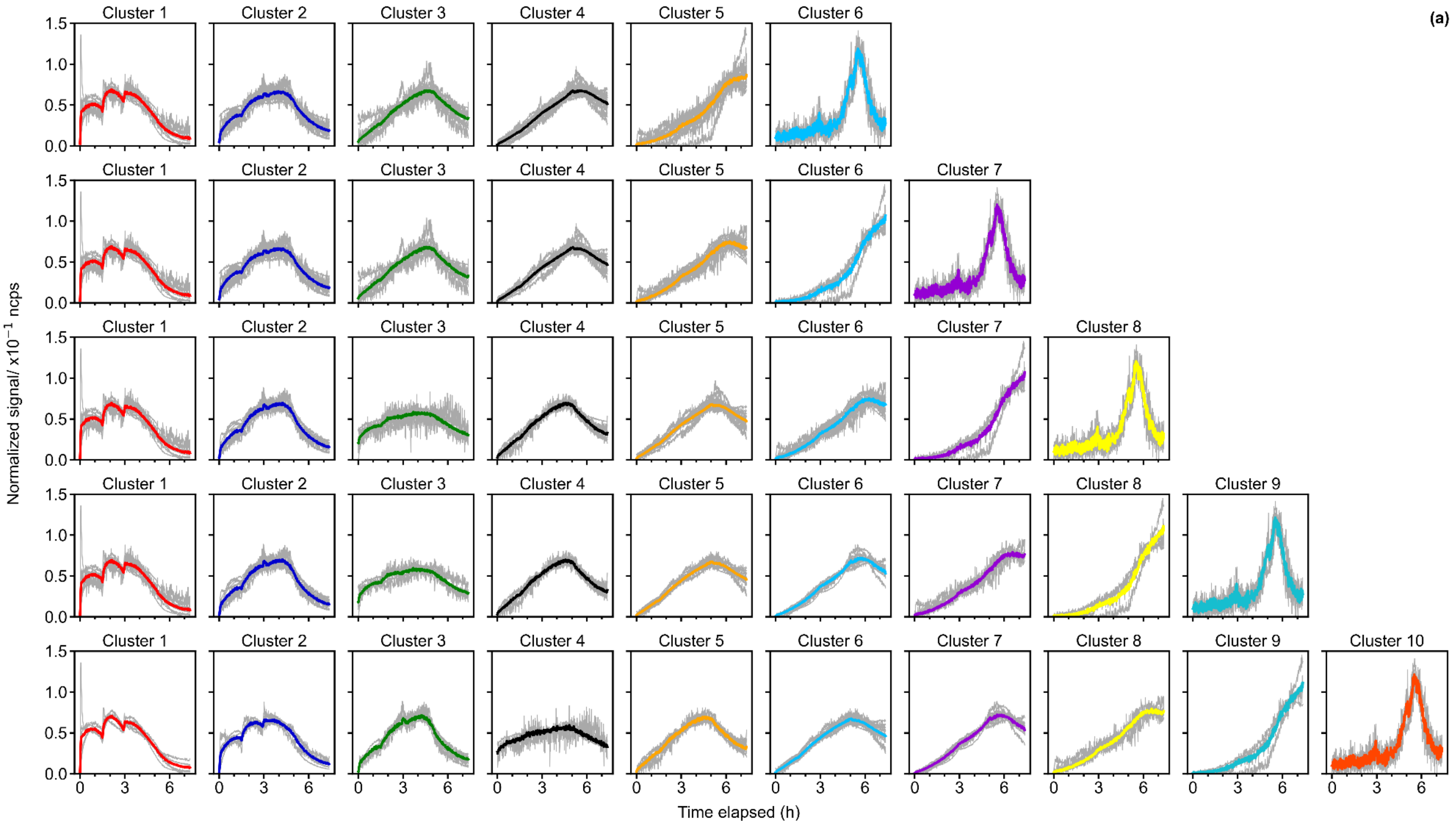
Figure S1. Concentrations of trace gases (NO_x, NO_y, and isoprene) and conditions of the chamber experiment selected for FCM analysis in this study. Adapted from Wu et al. (2021).

Fig S2: subscripts in O3, NO2, NO3

Response: Done. The caption of Fig. S2 has been updated accordingly.

Fig S4: same comments as main text figure

Response: Done. Figure S4 has been updated according to referee's comments, as shown in the following:


(a)
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
1.5
1.0
0.5
0.0
0
3
6
Normalized signal/ x10^{-1} ncps
Time elapsed (h)

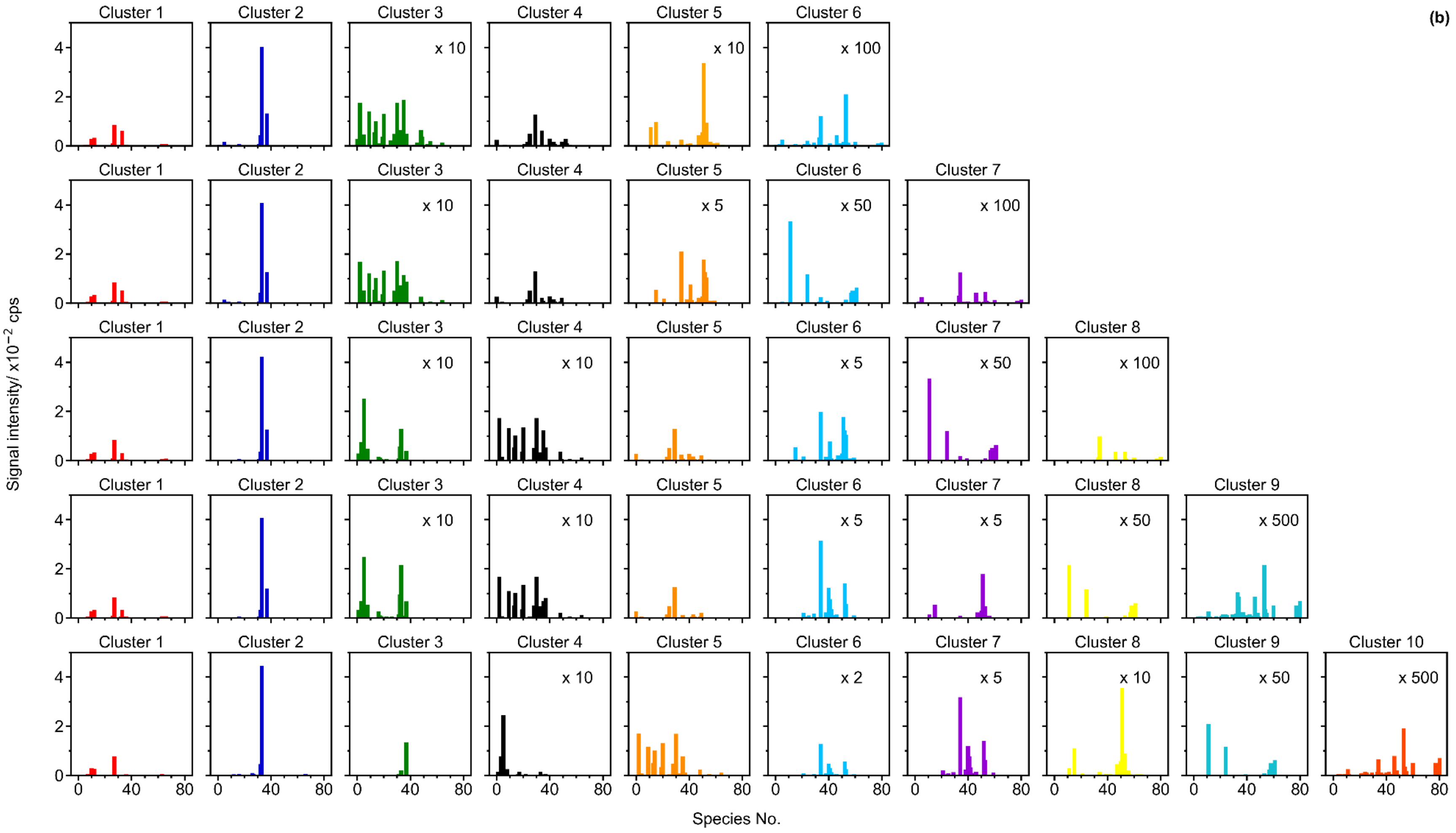


Figure S4. Fuzzy c-means clustering results of chamber data with 7-10 clusters. Time series (a) and profiles (b) of clusters for each solution. The cluster centers are shown as colored thick lines, and species with the membership degree larger than 0.5 to the cluster are illustrated as thin lines in gray. The species number in panel (b) corresponds to species listed in Fig. S7 (in order of molecular mass).

References

Huang, M., Xia, Z., Wang, H., Zeng, Q., and Wang, Q.: The range of the value for the fuzzifier of the fuzzy c-means algorithm, Pattern Recognition Letters, 33, 2280-2284, 2012.

Hwang, C. and Rhee, F. C.-H.: Uncertain fuzzy clustering: Interval type-2 fuzzy approach to c-means, IEEE Transactions on fuzzy systems, 15, 107-120, 2007.

Schwämmle, V. and Jensen, O. N.: A simple and fast method to determine the parameters for fuzzy c–means cluster analysis, Bioinformatics, 26, 2841-2848, 2010.

Yu, J., Cheng, Q., and Huang, H.: Analysis of the weighting exponent in the FCM, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34, 634-639, 2004.

Zhou, K., Fu, C., and Yang, S.: Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation, Science China Information Sciences, 57, 1-8, 2014.