Supplement of Atmos. Meas. Tech., 17, 1811–1835, 2024
https://doi.org/10.5194/amt-17-1811-2024-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

Application of fuzzy *c*-means clustering for analysis of chemical ionization mass spectra: insights into the gas phase chemistry of NO_3-initiated oxidation of isoprene

Rongrong Wu et al.

Correspondence to: Thomas F. Mentel (t.mentel@fz-juelich.de)

S1 Fuzzy clustering validity indices

Six fuzzy validity indices were used to determine the appropriate number of clusters, include Sum of within-cluster variance (V_{SWCV}), Fakuyama-Sugeno index (V_{FS}, Fukuyama, 1989), Xie-Beni index (V_{XB}, Xie and Beni, 1991), Kwon index (V_{Kwon}, Kwon, 1998), Bouguessa-Wang-Sun index (V_{BWS}, Bouguessa et al., 2006), and Fuzzy Silhouette (FS, Campello and Hruschka, 2006). Their definitions and notes for applications are described in this section.

(1) Sum of within-cluster variation (V_{SWCV}). The basic idea of clustering is to sort clusters so that the sum of within-cluster variation is minimized, and this is used as the objective function $J_m(U,V)$ in fuzzy c-means clustering, as given by Eq. S1. The sum of within-cluster squared distance measures the compactness of clustering, and the "elbow" point of the curve of V_{SWCV} as a function of numbers of clusters is generally considered as an indicator of the optimal number of clusters (Campello and Hruschka, 2006).

$$V_{SWCV} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d(x_j, v_i)^2 \tag{S1}$$

where x_j and v_i denote the j^{th} object in the dataset and the i^{th} cluster center, respectively, m is the fuzzifier, u_{ij} is the membership degree of x_j to the i^{th} cluster, and $d(x_j, v_i)$ represents the distance between the object x_j and the i$^{\text{th}}$ cluster center v_i.

The elbow point is where the V_{SWCV} stops to drop as rapidly as before, namely the point of maximum curvature. In this study, the KneedLocator function of Kneed package in Python was used to find the elbow point.

(2) Fukuyama-Sugeno index (V_{FS}). The Fakuyama-Sugeno index combines the membership degree and the geometrical property of the dataset to evaluate a partition (Bouguessa and Wang, 2004). It evaluates the quality of a clustering solution by measuring the discrepancy between compactness and separation of clusters, as formulated by Eq. S2. Obviously, smaller V_{FS} indicates better performance of clustering.

$$V_{FS} = Compact(c) - Separate_{sum}(c) \tag{S2}$$

where the compactness is defined by the sum of within-cluster squared distance, as given by Eq. S3:

$$Compact(c) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} d\left(x_j, v_i\right)^2 \tag{S3}$$

where x_j and v_i denote the j^{th} object in the dataset and the i^{th} cluster center, respectively, m is the fuzzifier, u_{ij} is the membership degree of x_j to the i^{th} cluster, and $d(x_j, v_i)$ represents the distance between the object x_j and the ith cluster center v_i.

And the separation of partition is measured by the sum of squared distances between each cluster center and the mean of all cluster centers, as given by Eq. S4:

$$Separate_{sum}(c) = \sum_{i=1}^{c}(\sum_{j=1}^{n} u_{ij}^{m})d(v_i, \bar{v})^2 \tag{S4}$$

where $\bar{v} = \frac{1}{c}\sum_{i=1}^{c} v_i$, and $d(v_i, \bar{v})$ represents the distance between the ith cluster center v_i and the mean of all cluster centers $\bar{v}$.

(3) Xie-Beni index (V_{XB}). Xie-Beni index is a popular fuzzy clustering validity measure proposed by Xie and Beni (1991). It is defined as the ratio of compactness and separation as shown in Eq. S5, where the sum of within-cluster squared distance divided by the total number of objects in the numerator, represents the compactness of the partition, and the minimum squared distance of cluster centers in the denominator represents the separation. The smaller the numerator, the more compact the clusters are, whereas the larger the denominator, the more dispersed the clusters are from each other. As a consequence, the smaller V_{XB}, the better the partition.

$$V_{XB} = \frac{\frac{1}{n}Compact(c)}{Separate_{min}(c)} \tag{S5}$$

where n is the total number of objects in the data set, the compactness of the partition, $Compact(c)$, is defined by the sum of within-cluster squared distance, as given by Eq. S3, and the separation of partition, $Separate_{min}(c)$, is measured by the minimum squared distance between cluster centers, as calculated by Eq. S6:

$$Separate_{min}(c) = \min_{k \neq i} d(v_k, v_i)^2 \tag{S6}$$

where $d(v_k, v_i)$is the distance between the kth cluster center v_k and the ith cluster center v_i $(k \neq i$.

(4) Kwon index (V_{kwon}). When c approaches n, the value of V_{XB} decreases monotonically to 0 and will lose robustness in determining the optimal number of clusters. To overcome this drawback, Kwon (1998) revised V_{XB} and proposed the Kwon index, as defined in Eq. S7. The second item in the numerator is a penalty function, which represents the average squared distance of cluster centers to the overall mean of the data set and can eliminate its monotonous decreasing tendency when the number of clusters is close to n. Similar to V_{XB}, the smaller V_{kwon}, the better the clustering quality.

$$V_{Kwon} = \frac{Compact(c)+Penalty(c)}{Separate_{min}(c)} \tag{S7}$$

where $Compact(c)$ and $Separate_{min}(c)$ represent the compactness and separation of the partition, which are calculated by Eq. S3 and Eq. S6, respectively, and $Penalty(c)$ is a penalty function defined by Eq. S8:

$$Penalty(c) = \ (1/c)\sum_{i=1}^{c} d(v_i, \bar{x})^2 \tag{S8}$$

where $\bar{x}$ denotes the overall mean of the data set, that is $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, and $d(v_i, \bar{x})$ is the distance between the ith cluster center v_i and $\bar{x}$.

(5) Bouguessa-Wang-Sun index (V_{BWS}). To better deal with overlapped clusters that differ in geometric shape, Bouguessa et al. (2006) proposed a new validity index, as formulated in Eq. S9, and hereafter called Bouguessa-Wang-Sun index in this study. Similar to V_{XB} and V_{Kwon}, V_{BWS} is also based on the concept of using the ratio of separation and compactness, but the definitions for compactness and separation are modified. By making use of the fuzzy covariance matrix as a measure of compactness, V_{BWS} takes the variations of cluster shape, density and orientation into account and was proved to performe well for heavily overlapping clusters (Bouguessa and Wang, 2004; Bouguessa et al., 2006). According to its definition, a larger value of V_{BWS} indicates a better fuzzy partition.

$$V_{BWS} = \frac{Sep(c)}{Comp(c)} \tag{S9}$$

In the equation, $Sep(c)$ represents fuzzy separation, as defined in Eq. S10, and $\boldsymbol{S_B}$ is the between-cluster fuzzy matrix given by Eq. S11. The larger $Sep(c)$, the better separation between clusters.

$$Sep(c) = trace(S_B) \tag{S10}$$

$$S_B = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}(v_i - \bar{v})\,(v_i - \bar{v})^T \tag{S11}$$

$Comp(c)$ in Eq. S9 represents the overall compactness of fuzzy clustering, as given by Eq. S12. The smaller $Comp(c)$, the more compact within each cluster.

$$Comp(c) = \sum_{i=1}^{c} trace(\sum_i) \tag{S12}$$

where $\sum_i$ is the fuzzy covariance matrix as defined by:

$$\sum_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} (x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{S13}$$

(6) Fuzzy Silhouette (FS). The silhouette score (s_j, as defined in Eq. S14) was first proposed by Rousseeuw (1987) and can be used to measure how close an object is to the cluster center it belongs compared to other clusters. The average silhouette score of all objects, CS, as given by Eq. S15, are frequently used to assess the quality of clustering solutions. The silhouette score was originally adopted to evaluate hard or non-fuzzy clustering solutions and did not consider the fuzzy partition matrix in the calculation. Consequently, CS might be inadequate to discriminate fuzzy clusters since it ignores the information contained in the fuzzy partition matrix which reveal the overlap degrees of clusters.

$$s_j = \frac{b_j - a_j}{max\{a_j, b_j\}} \tag{S14}$$

$$CS = \frac{1}{n} \sum_{j=1}^{n} s_j \tag{S15}$$

where a_j is the average distance of object x_j (belonging to cluster p) to all other objects in the same cluster. Let d_j be the average distance of object x_j to all objects belonging to another cluster r ($r \neq i$), then b_j is the minimum of d_j, which represents the average distance of object j to its closet neighboring cluster.

To extend the silhouette score to fuzzy partition and make explicit use of the fuzzy partition matrix, Campello and Hruschka (2006) proposed *Fuzzy Silhouette* (FS), as given by Eq. S16. Instead of weighing each individual silhouette equally, FS stresses the importance of objects lying in the vicinity of cluster centers while reducing the importance of objects located in the boundary region (whose membership degrees to different clusters are similar or identical).

$$FS = \frac{\sum_{j=1}^{n} (u_{pj} - u_{qj})^{\alpha} s_j}{\sum_{j=1}^{n} (u_{pj} - u_{qj})^{\alpha}} \tag{S16}$$

where s_j in the average silhouette score of object x_j, u_{pj} and u_{qj} are the first and second largest coefficient of x_j in the fuzzy partition matrix, respectively, and α is a weight coefficient and set to be 1 as default in this study (Campello and Hruschka, 2006).

In a hard partition, each object is exclusively partitioned to one cluster, and it is easier to determine the intra- (within-cluster) and inter- (between-cluster) distances. With regard to a fuzzy partition, however, an object could belong to multiple clusters simultaneously, and its affiliation to each cluster is measured by the membership degree. In order to determine the intra- and inter-distance of an object in a fuzzy partition, the original definition of silhouette is reformed by introducing a concept of intra-inter scores. The intra-score matrix is defined by

$$IntraDist_i = \left[intra_i\left(d_{jk}\right)\right], \quad 1 \leq j \neq k \leq n, 1 \leq i \leq c \tag{S17}$$

where $intra_i\left(d_{jk}\right) = \left(u_{ij} \wedge u_{ik}\right)$.

And the inter-score matrix is given by

$$InterDist_{ir} = \left[inter_{ir}\left(d_{jk}\right)\right], \quad 1 \leq j \neq k \leq n, 1 \leq i < r \leq c \tag{S18}$$

where $inter_{ir}\left(d_{jk}\right) = \left(u_{ij} \wedge u_{rk}\right) \vee \left(u_{rj} \wedge u_{ik}\right)$.

u_{ij} and u_{ik} are the membership degree of object x_j and x_k to cluster i, and u_{rj} and u_{rk} are the membership degree of object x_j and x_k to cluster *r*, respectively.

With the intra- and inter-distance scores defined above, we can calculate the intra-distance a_j and inter-distance b_j follow the equations proposed by Rawashdeh and Ralescu (2012), as shown by Eq. S19 amd Eq. S20, respectively:

$$a_j = \min\left\{\frac{\sum_{k=1}^{n} IntraDist_i(j,k) d\left(x_j, x_k\right)}{\sum_{k=1}^{n} IntraDist_i(j,k)}\right\}, \quad 1 \leq j \neq k \leq n, 1 \leq i \leq c \tag{S19}$$

$$b_j = \min\left\{\frac{\sum_{k=1}^{n} InterDist_{ir}(j,k) d\left(x_j, x_k\right)}{\sum_{k=1}^{n} InterDist_{ir}(j,k)}\right\}, \quad 1 \leq j \neq k \leq n, 1 \leq i < r \leq c \tag{S20}$$

where $IntraDist_i(j,k)$ and $InterDist_{ir}(j,k)$ are the intra-, and inter-distance score of the object x_j, respectively, as defined in Eq. S17 and Eq. S18, and $d(x_j, x_k)$ represents the distance between the object x_j and x_k.

The silhouette score falls in the range from -1 to +1, with a value approaching +1 indicating that the object is correctly assigned, whereas with a value close to -1 indicating that the object is misclustered (better to sort it to a neighboring cluster than to current cluster). An s_j close to 0 implies that the object lies in the boundary region (between clusters) and thus it is unclear to which cluster it belongs. The average cluster silhouette score can tell if the cluster is appropriately configurated or not. The larger the average cluster silhouette score, the clearer the cluster. The overall average silhouette score of all objects in the dataset can be used as a measure of clustering quality. Further, it can be used to find the appropriate number of clusters. When plotting the overall silhouette score as a function of cluster number, the maximum point of the curve indicates the optimal value of c, where the clustering solution has a minimum intra-cluster distance (a_j) and a maximum inter-cluster distance (b_j).

Table S1. Possible permutation scheme for 2N- (grey), 3N- (blue) and 4N-dimers (orange) formed through RO_2 + $R'O_2$ reactions. Second-generation species are outlined in blue. And molecules detected by Br^- CIMS are shown in bold.

$C_5H_8NO_x$ **$C_5H_8NO_x$**	**O_5**	**O_6**	**O_7**	**O_8**	**O_9**	**O_{10}**	**O_{11}**
O_5	**$C_{10}H_{16}N_2O_8$**						
O_6	**$C_{10}H_{16}N_2O_9$**	**$C_{10}H_{16}N_2O_{10}$**					
O_7	**$C_{10}H_{16}N_2O_{10}$**	**$C_{10}H_{16}N_2O_{11}$**	**$C_{10}H_{16}N_2O_{12}$**				
O_8	**$C_{10}H_{16}N_2O_{11}$**	**$C_{10}H_{16}N_2O_{12}$**	**$C_{10}H_{16}N_2O_{13}$**	$C_{10}H_{16}N_2O_{14}$			
O_9	**$C_{10}H_{16}N_2O_{12}$**	**$C_{10}H_{16}N_2O_{13}$**	$C_{10}H_{16}N_2O_{14}$	$C_{10}H_{16}N_2O_{15}$	$C_{10}H_{16}N_2O_{16}$		
O_{10}	**$C_{10}H_{16}N_2O_{13}$**	$C_{10}H_{16}N_2O_{14}$	$C_{10}H_{16}N_2O_{15}$	$C_{10}H_{16}N_2O_{16}$	$C_{10}H_{16}N_2O_{17}$	$C_{10}H_{16}N_2O_{18}$	
O_{11}	$C_{10}H_{16}N_2O_{14}$	$C_{10}H_{16}N_2O_{15}$	$C_{10}H_{16}N_2O_{16}$	$C_{10}H_{16}N_2O_{17}$	$C_{10}H_{16}N_2O_{18}$	$C_{10}H_{16}N_2O_{19}$	$C_{10}H_{16}N_2O_{20}$

$C_5H_8NO_x$ **$C_5H_9N_2O_y$**	O_5	O_6	O_7	O_8	O_9	O_{10}	O_{11}
O_9	**$C_{10}H_{17}N_3O_{12}$**	**$C_{10}H_{17}N_3O_{13}$**	**$C_{10}H_{17}N_3O_{14}$**	**$C_{10}H_{17}N_3O_{15}$**	**$C_{10}H_{17}N_3O_{16}$**	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**
O_{10}	**$C_{10}H_{17}N_3O_{13}$**	**$C_{10}H_{17}N_3O_{14}$**	**$C_{10}H_{17}N_3O_{15}$**	**$C_{10}H_{17}N_3O_{16}$**	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$
O_{11}	**$C_{10}H_{17}N_3O_{14}$**	**$C_{10}H_{17}N_3O_{15}$**	**$C_{10}H_{17}N_3O_{16}$**	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$
O_{12}	**$C_{10}H_{17}N_3O_{15}$**	**$C_{10}H_{17}N_3O_{16}$**	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$	$C_{10}H_{17}N_3O_{21}$
O_{13}	**$C_{10}H_{17}N_3O_{16}$**	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$	$C_{10}H_{17}N_3O_{21}$	$C_{10}H_{17}N_3O_{22}$
O_{14}	$C_{10}H_{17}N_3O_{17}$	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$	$C_{10}H_{17}N_3O_{21}$	$C_{10}H_{17}N_3O_{22}$	$C_{10}H_{17}N_3O_{23}$
O_{15}	**$C_{10}H_{17}N_3O_{18}$**	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$	$C_{10}H_{17}N_3O_{21}$	$C_{10}H_{17}N_3O_{22}$	$C_{10}H_{17}N_3O_{23}$	$C_{10}H_{17}N_3O_{24}$
O_{16}	$C_{10}H_{17}N_3O_{19}$	$C_{10}H_{17}N_3O_{20}$	$C_{10}H_{17}N_3O_{21}$	$C_{10}H_{17}N_3O_{22}$	$C_{10}H_{17}N_3O_{23}$	$C_{10}H_{17}N_3O_{24}$	$C_{10}H_{17}N_3O_{25}$

$C_5H_9N_2O_y$ **$C_5H_9N_2O_y$**	O_9	O_{10}	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}
O_9	**$C_{10}H_{18}N_4O_{16}$**						
O_{10}	$C_{10}H_{18}N_4O_{17}$	$C_{10}H_{18}N_4O_{18}$					
O_{11}	$C_{10}H_{18}N_4O_{18}$	$C_{10}H_{18}N_4O_{19}$	$C_{10}H_{18}N_4O_{20}$				
O_{12}	$C_{10}H_{18}N_4O_{19}$	$C_{10}H_{18}N_4O_{20}$	$C_{10}H_{18}N_4O_{21}$	$C_{10}H_{18}N_4O_{22}$			
O_{13}	$C_{10}H_{18}N_4O_{20}$	$C_{10}H_{18}N_4O_{21}$	$C_{10}H_{18}N_4O_{22}$	$C_{10}H_{18}N_4O_{23}$	$C_{10}H_{18}N_4O_{24}$		
O_{14}	$C_{10}H_{18}N_4O_{21}$	$C_{10}H_{18}N_4O_{22}$	$C_{10}H_{18}N_4O_{23}$	$C_{10}H_{18}N_4O_{24}$	$C_{10}H_{18}N_4O_{25}$	$C_{10}H_{18}N_4O_{26}$	
O_{15}	$C_{10}H_{18}N_4O_{22}$	$C_{10}H_{18}N_4O_{23}$	$C_{10}H_{18}N_4O_{24}$	$C_{10}H_{18}N_4O_{25}$	$C_{10}H_{18}N_4O_{26}$	$C_{10}H_{18}N_4O_{27}$	$C_{10}H_{18}N_4O_{28}$
O_{16}	$C_{10}H_{18}N_4O_{23}$	$C_{10}H_{18}N_4O_{24}$	$C_{10}H_{18}N_4O_{25}$	$C_{10}H_{18}N_4O_{26}$	$C_{10}H_{18}N_4O_{27}$	$C_{10}H_{18}N_4O_{28}$	$C_{10}H_{18}N_4O_{29}$

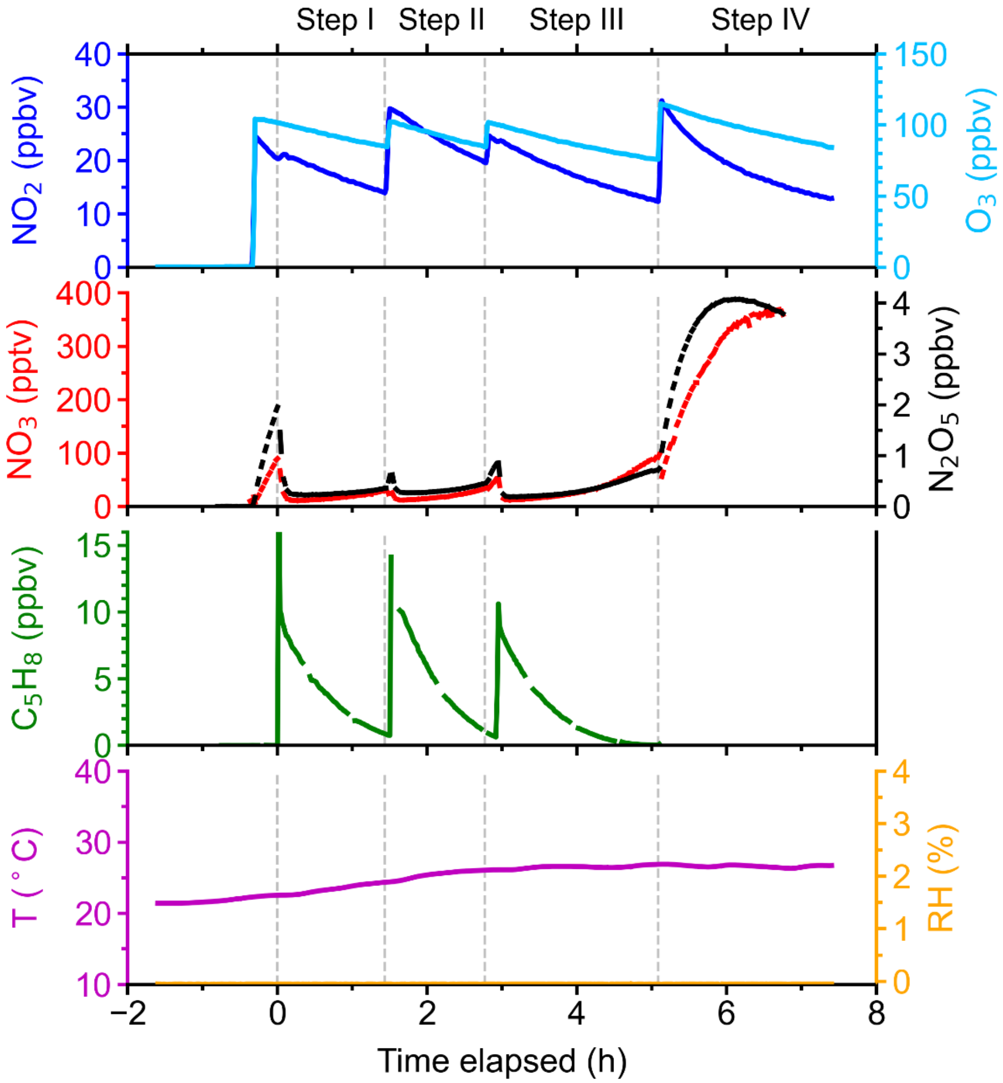


Figure S1. Concentrations of trace gases (NO_x, NO_y, and isoprene) and conditions of the chamber experiment selected for FCM analysis in this study. Adapted from Wu et al. (2021).

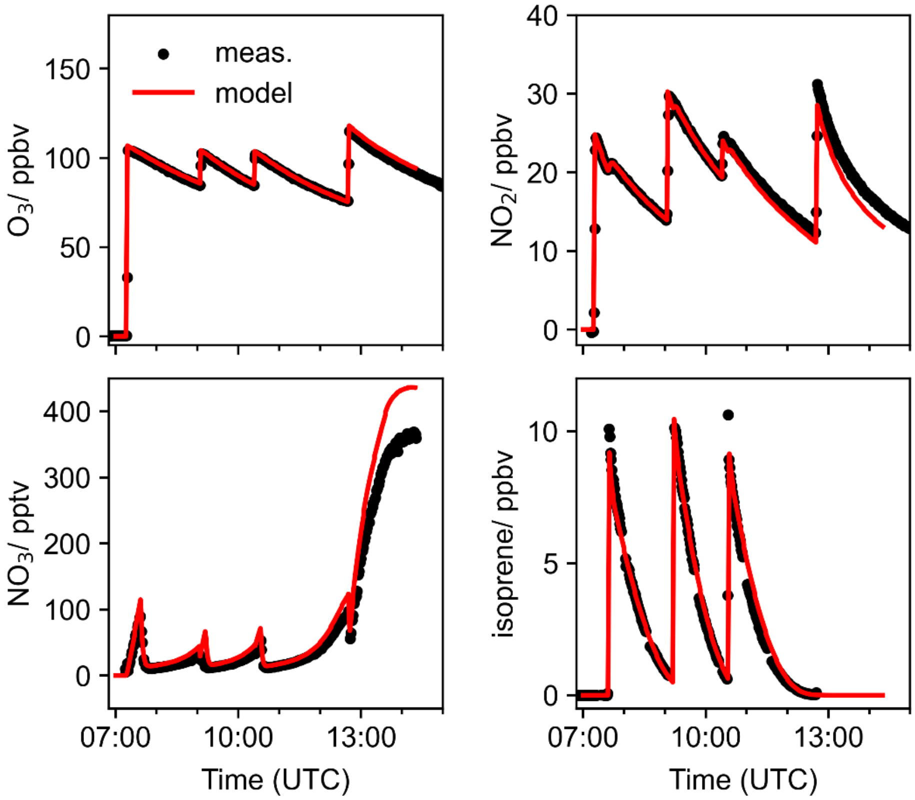


Figure S2. Measured and simulated concentrations of O_3, NO_2, NO_3, and isoprene in the chamber experiment of isoprene oxidation by NO_3. Simulation results are from a box model with using the gas-phase chemistry mechanism of isoprene + NO_3 from MCM v3.3.1.

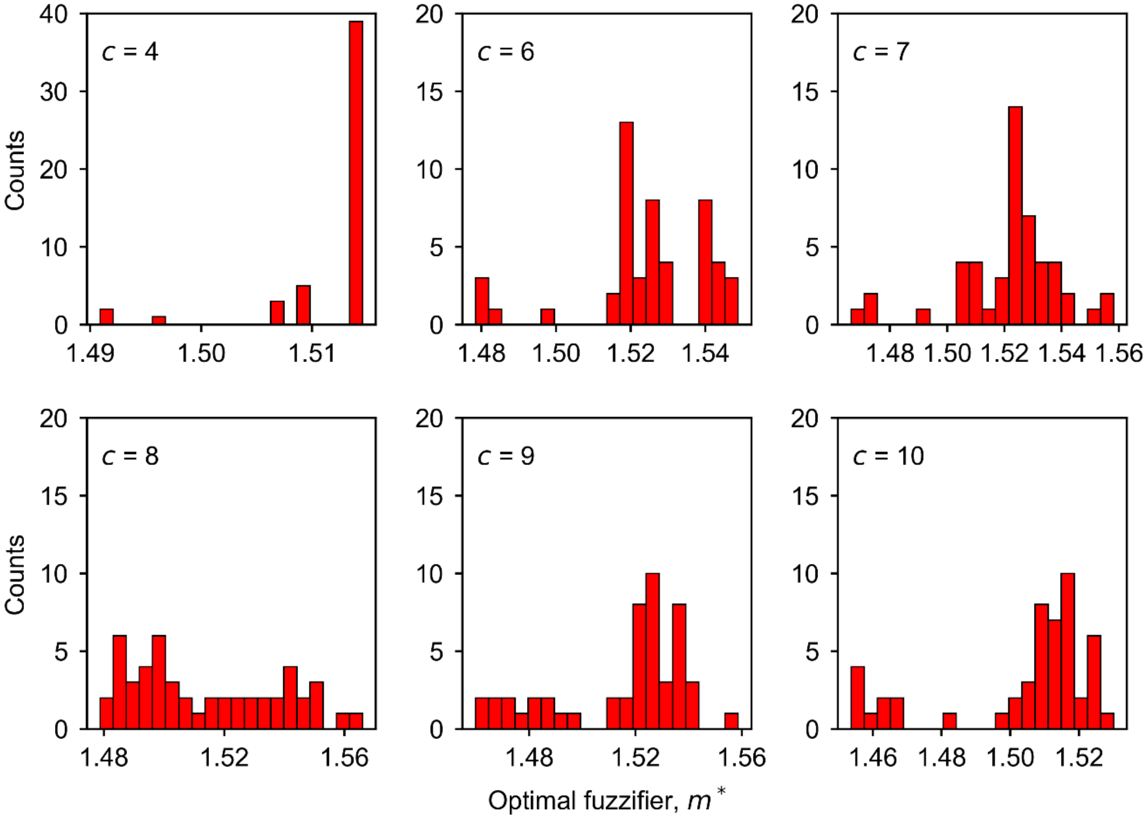
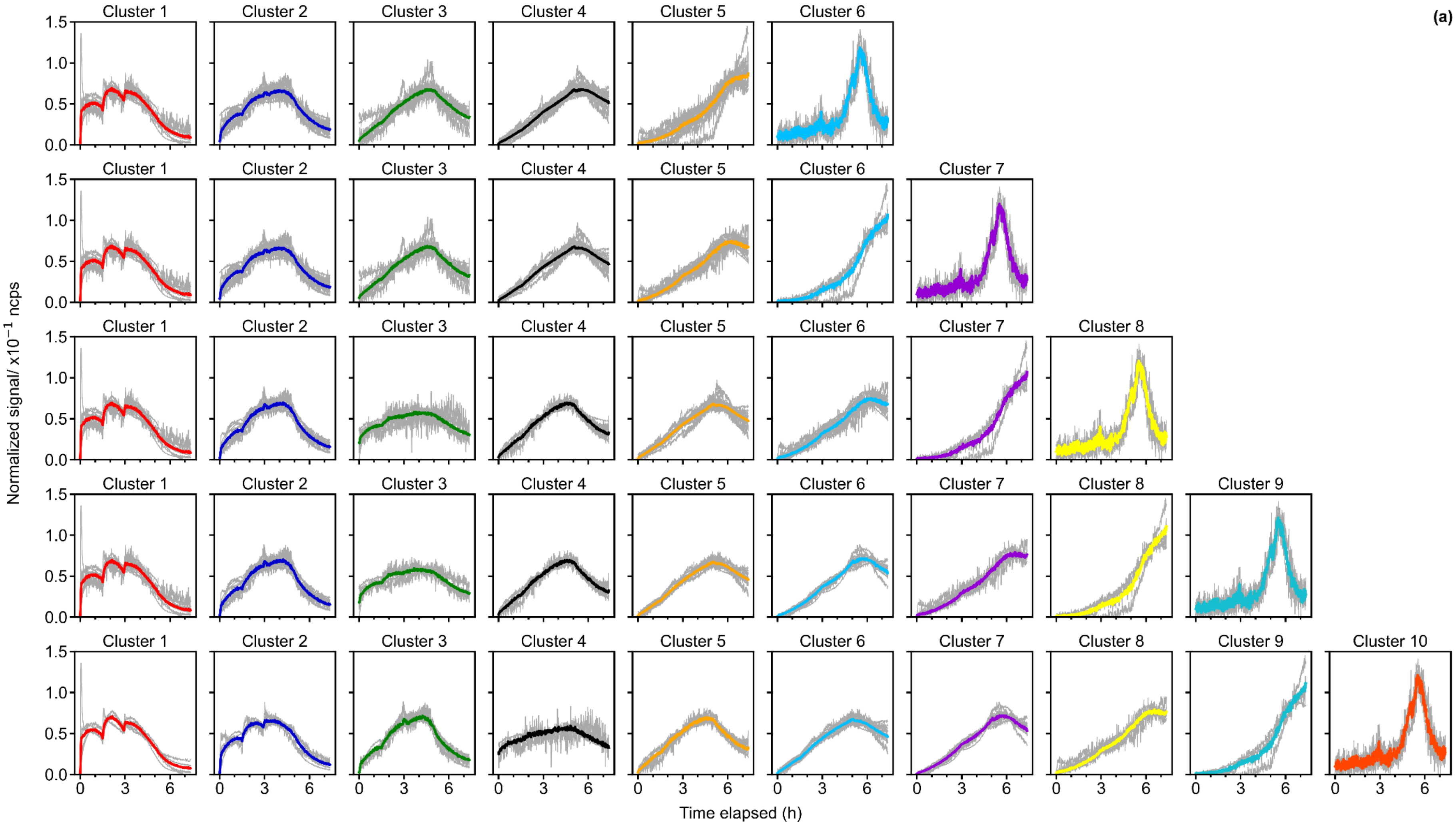


Figure S3. Distribution of the optimal value of fuzzifier (m^{*}) obtained from 50 repetitions.


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

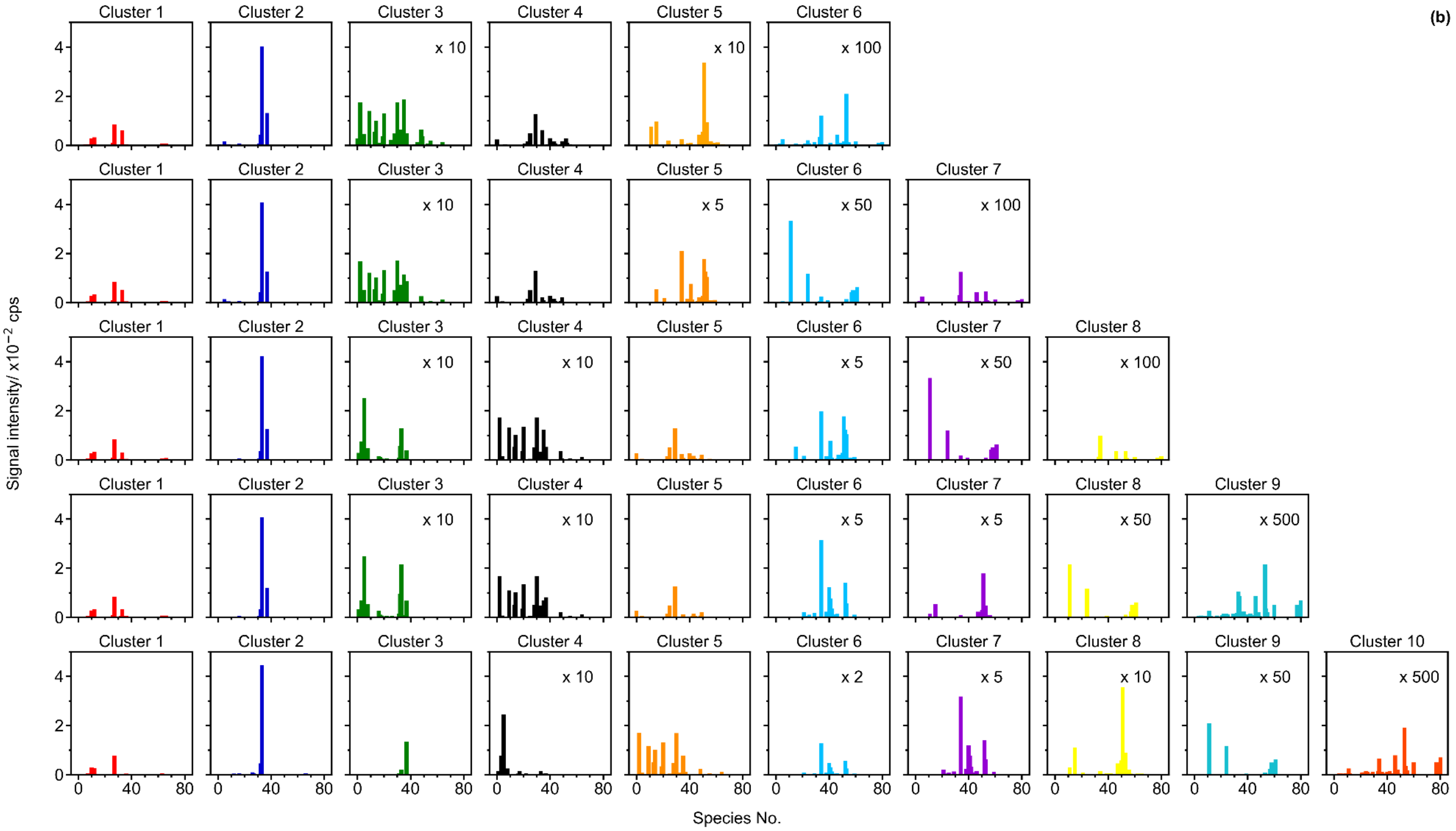


Figure S4. Fuzzy c-means clustering results of chamber data with 7-10 clusters. Time series (a) and profiles (b) of clusters for each solution. The cluster centers are shown as colored thick lines, and species with the membership degree larger than 0.5 to the cluster are illustrated as thin lines in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

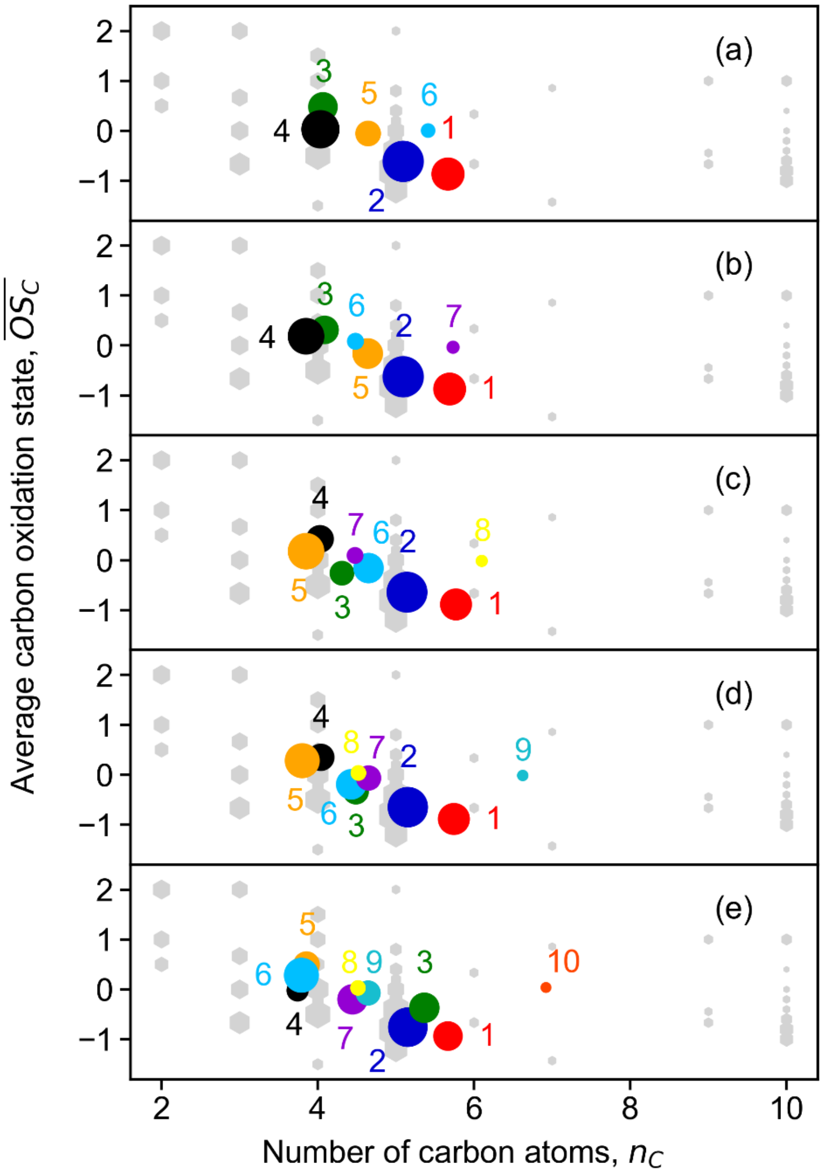


Figure S5. Average oxidation state ($\overline{OS_C}$) of FCM clusters of chamber data as a function of number of carbon atoms (n_C). Panel (a) to panel (e) show results for solutions with 6 to 10 clusters, respectively. Cluster centers are depicted by circles in different colors. The color scheme follows 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. Closed-shell products detected by Br- CIMS are shown as grey hexagons, and the marker area is proportional to the average intensity of species over the whole experiment.

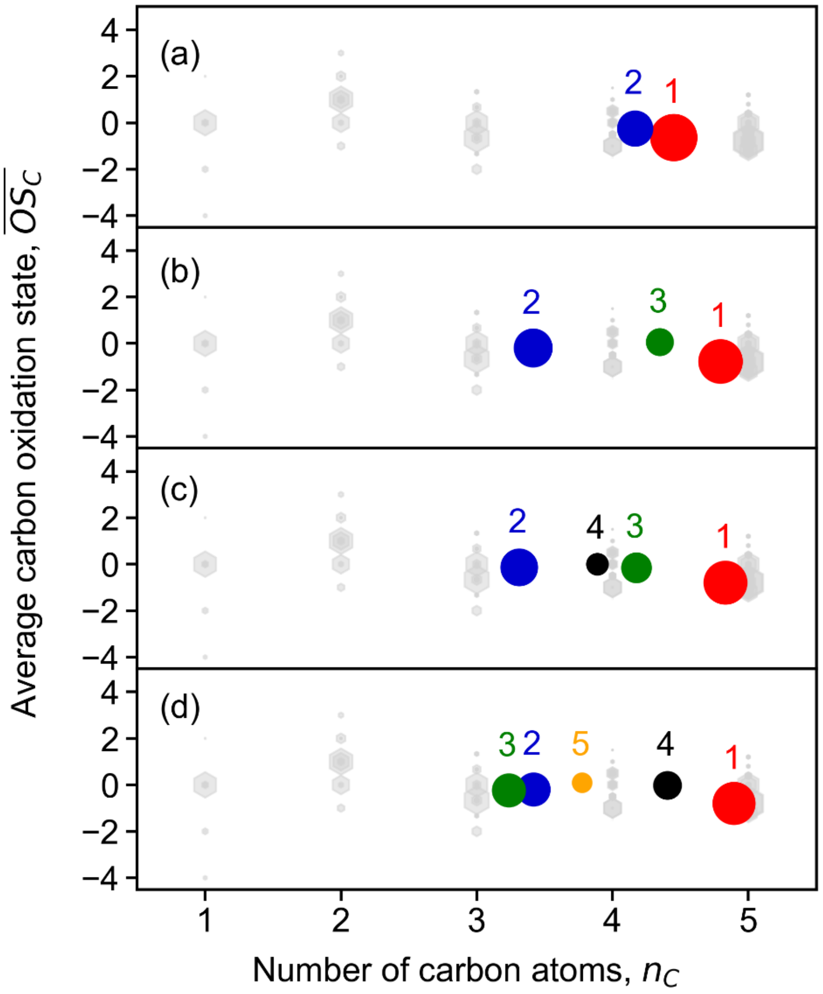


Figure S6. Average oxidation state ($\overline{OS_C}$) of FCM clusters of model data as a function of number of carbon atoms (n_C). Panel (a) to panel (d) show results for solutions with 2 to 5 clusters, respectively. Cluster centers are depicted by circles in different colors. The color scheme follows that in Fig. 4. The marker area of clusters is proportional to the sum of the average signal intensity of all species in the cluster weighted by their membership degrees. Closed-shell products detected by Br^- CIMS are shown as grey hexagons, and the marker area is proportional to the average intensity of species over the whole experiment.

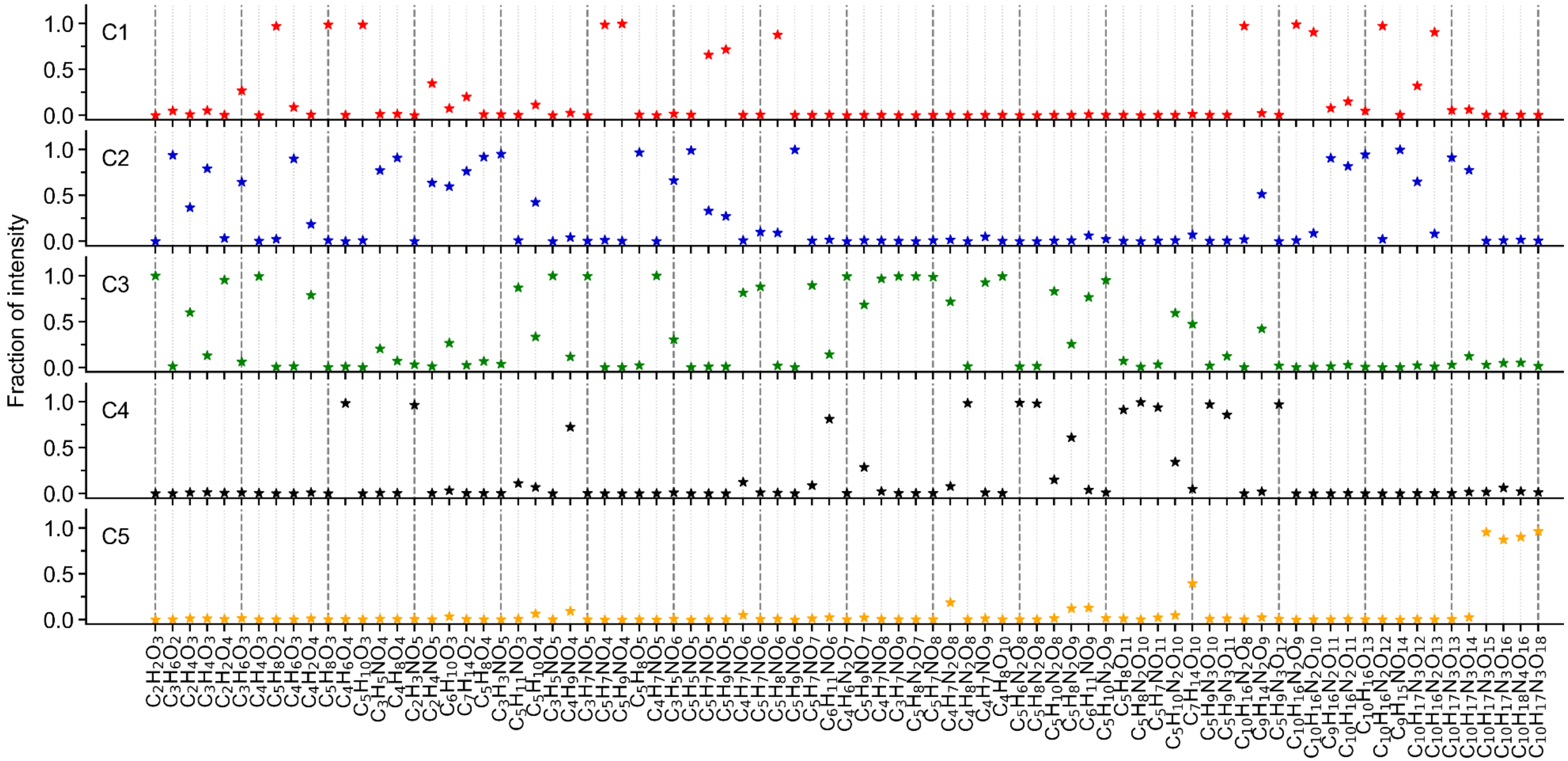


Figure S7. Cluster apportionment of species for the five-cluster solution. The sum of fractions of a compound in each cluster adds up to 1. Different clusters are distinguished by color, and the color scheme follows that in Fig. 4. Species are listed in the same order (in order of molecular mass) to those in Fig. 7.

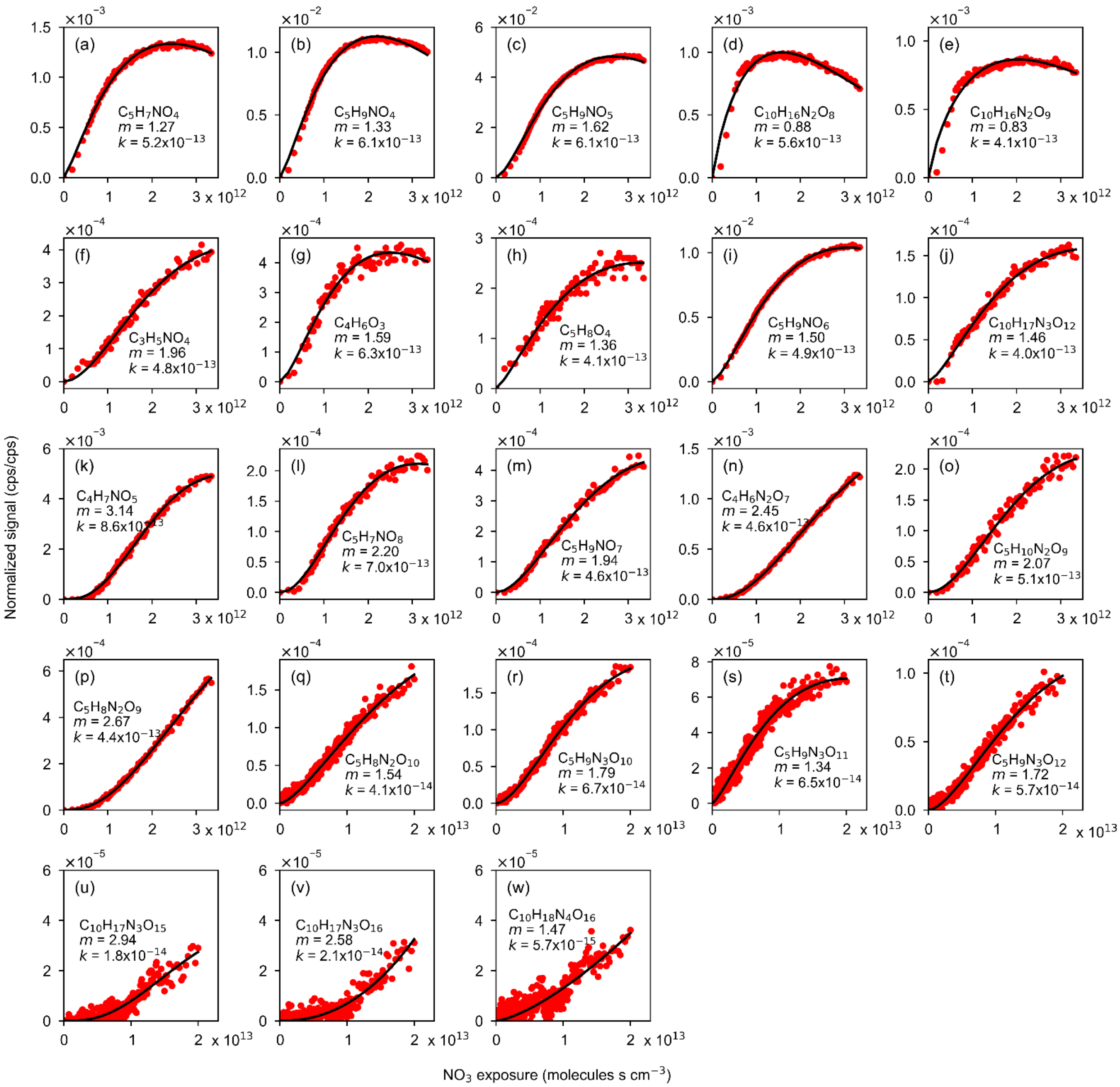
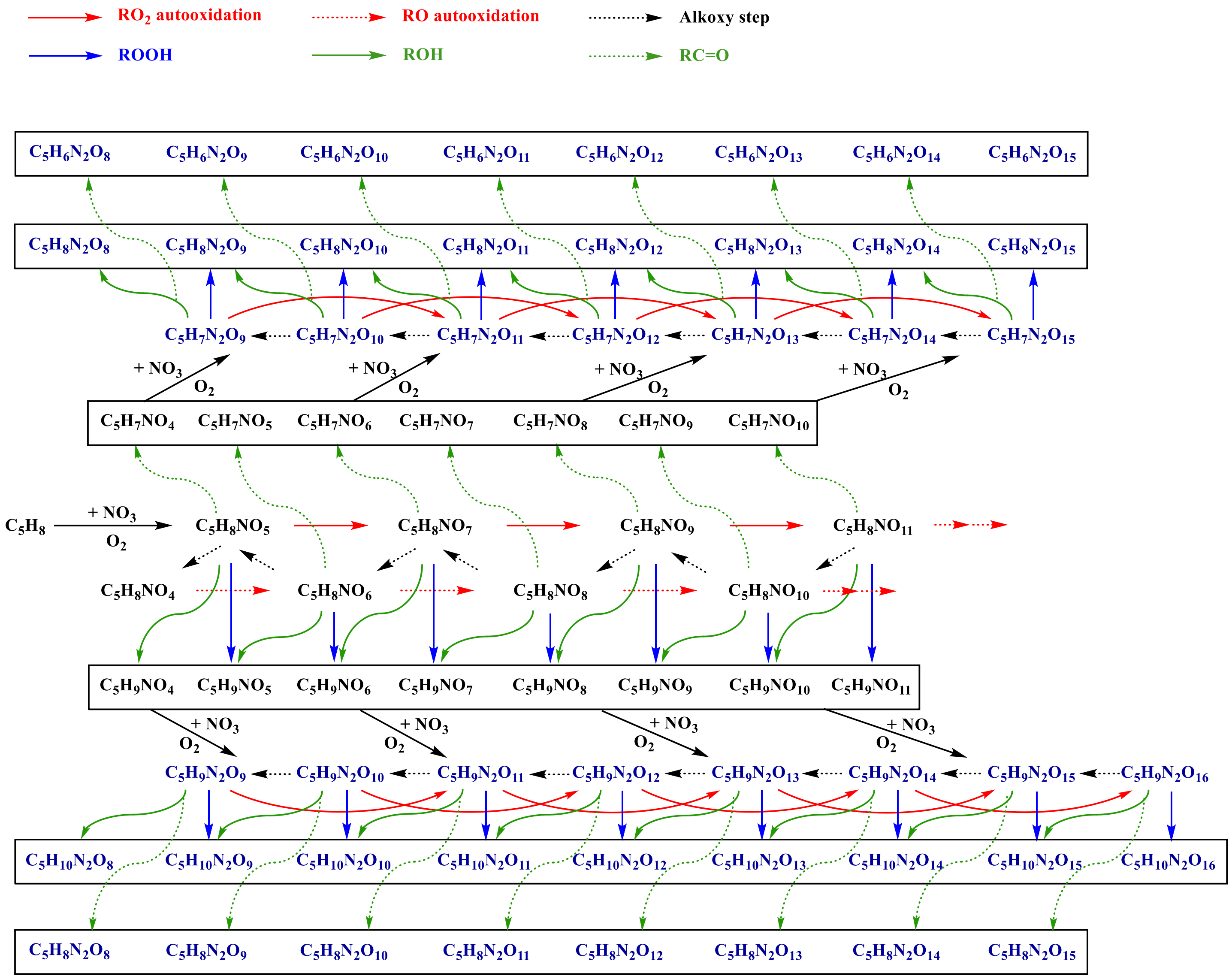
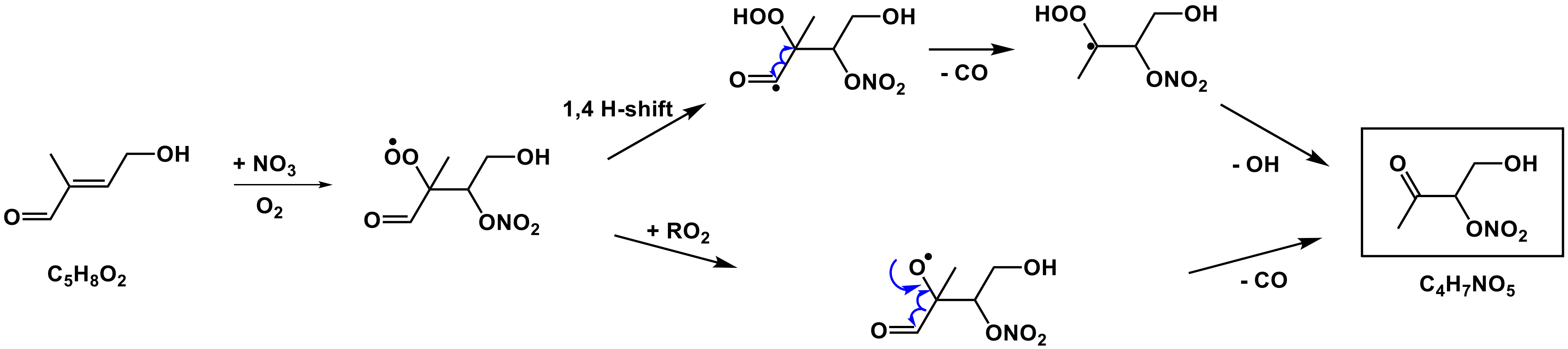


Figure S8. Representative species measured by Br$^-$-CIMS from isoprene + NO3 experiment (red) and the GKP fitting results (black).



Scheme S1. General reaction scheme of isoprene oxidation by NO_3. The first- and second-generation products are shown in black and blue, respectively. Closed-shell species are outlined in black boxes. Dimers are not shown in this scheme for simplicity.



Scheme S2. Proposed formation mechanism of $C_4H_7NO_5$ through further oxidation of the first-generation C_5 carbonyl compound. Adapted from Wu et al. (2021).

References

Bouguessa, M. and Wang, S.-R.: A new efficient validity index for fuzzy clustering, Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826), 1914-1919,

Bouguessa, M., Wang, S., and Sun, H.: An objective approach to cluster validation, Pattern Recognition Letters, 27, 1419-1430, 10.1016/j.patrec.2006.01.015, 2006.

Campello, R. J. G. B. and Hruschka, E. R.: A fuzzy extension of the silhouette width criterion for cluster analysis, Fuzzy Sets and Systems, 157, 2858-2875, 10.1016/j.fss.2006.07.006, 2006.

Fukuyama, Y.: A new method of choosing the number of clusters for the fuzzy c-mean method, Proc. 5th Fuzzy Syst. Symp., 1989, 247-250.

Kwon, S.-H.: Cluster validity index for fuzzy clustering, Electronics Letters, 34, 2176-2177, 1998.

Rousseeuw, P. J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of computational and applied mathematics, 20, 53-65, 1987.

Rawashdeh, M. and Ralescu, A. L.: Fuzzy Cluster Validity with Generalized Silhouettes, Midwest Artificial Intelligence and Cognitive Science Conference, 2012.

Xie, X. L. and Beni, G.: A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis & Machine Intelligence, 13, 841-847, 1991.