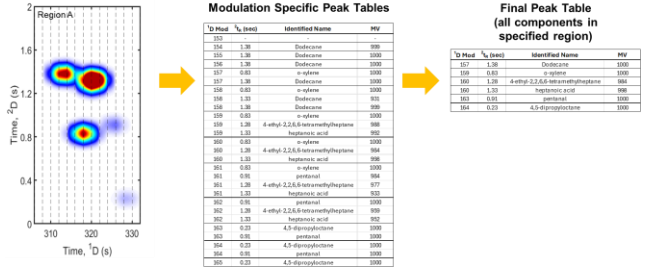
Developing 2D mzCompare for single GC×GC-TOFMS chromatograms: Substantial resolution enhancement in the context of statistical overlap theory

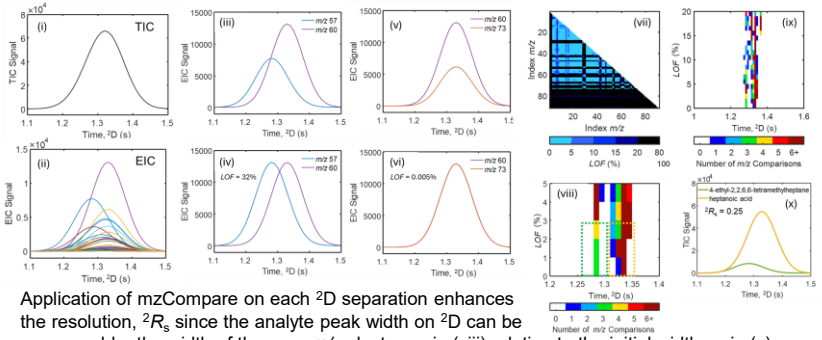
Wenjing Ma, Caitlin N. Cain, Robert E. Synovec
Department of Chemistry, University of Washington, Seattle, WA

Overview

Accurate identification of all detectable analyte components in a single comprehensive two-dimensional (2D) gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) chromatogram is a fundamental interest in the field. Herein, we developed a new algorithmic software approach called 2D mzCompare to generate accurate peak tables for GC×GC-TOFMS. Extending from our original method for one-dimensional GC-MS data, the 2D mzCompare algorithm discovers selective mass channels (*m/z*) for each analyte to resolve overlapping peaks and improve analyte identification, leveraging the similarity in retention time and peak shape across *m/z* of the same analyte. The 2D mzCompare algorithm calculates the peak shape similarity between *m/z* at every modulation via lack-of-fit (*LOF*), followed by clustering and focusing steps, to generate a final peak table. To evaluate this software, we simulated realistic GC×GC-TOFMS data in the context of the statistical overlap theory (SOT), so the exact number and identities of analytes are known *a priori*. Utilizing an in-house mass spectrum library of similar compounds, GC×GC-TOFMS chromatograms were simulated with varying degrees of chromatographic saturation (α_{2D}). First, we provide a new algorithmic approach, 2D mzCompare, to resolve overlapped analytes in GC×GC-TOFMS data, and second, we validate the accuracy of the software performance using SOT.



1D mzCompare Algorithm

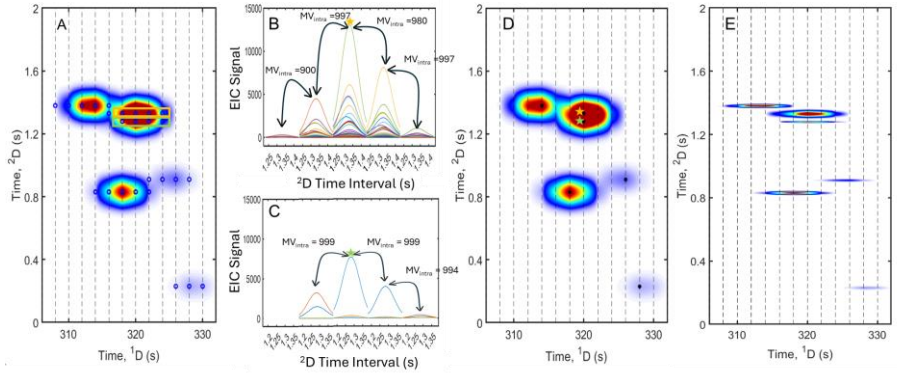


Application of mzCompare on each ^{2}D separation enhances the resolution, ^2R_s since the analyte peak width on ^{2}D can be expressed by the width of the pure *m/z* cluster as in (viii) relative to the initial width as in (x),

$$R_S = \frac{(t_{R_2} - t_{R_1})}{(w_{b_2} + w_{b_1})/2} \qquad R_{s,mzCompare} = R_S\left(\frac{w_{b_2} + w_{b_1}}{w_{b,cluster_2} + w_{b,cluster_1}}\right)$$

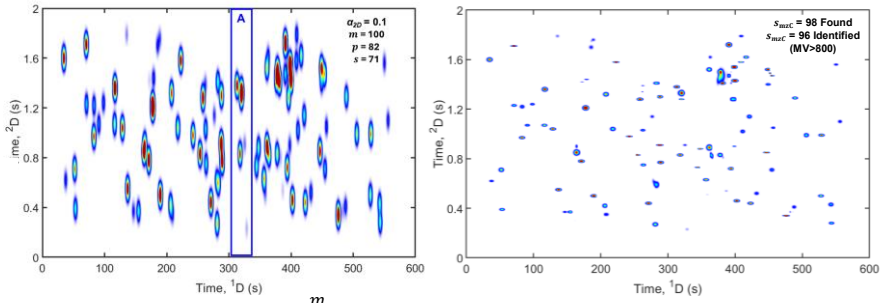
where $^2w_{b,cluster_2}$ and $^2w_{b,cluster_1}$ are pure *m/z* cluster ^{2}D "peak widths" in (viii).

2D mzCompare Algorithm: Clustering and Focusing



The ^2t_R **locations for each analyte** (**blue circles** in A) are obtained from applying mzCompare at each modulation. These are clustered across adjacent modulations via intra-mass spectrum comparison, and when the MV exceeds 800 and falls within the ±2-pixel cluster window, a **final singlet component location** is obtained (**black dots** in D).

2D mzCompare Results at α_{2D} = 0.1



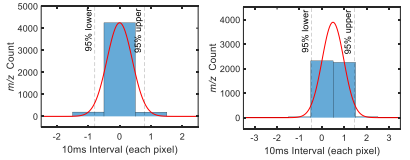
Saturation factor (α_{2D}) $\alpha_{2D} = \frac{m}{n_{c,2D}}$

m = number of analytes in simulation

$n_{c,2D}$ = ideal peak capacity of a 2D chromatogram

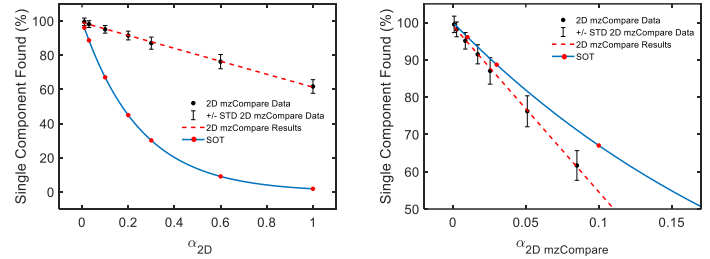
Number of peaks (*p*) $p = m\frac{4\alpha_{2D}e^{-4\alpha_{2D}}}{1 - e^{-4\alpha_{2D}}}$

Number of singlets (*s*) $s = me^{-4\alpha_{2D}}$



Cumulative distribution for pure *m/z* cluster plots from 100 singlets with *LOF* below 3%. The exponential decay peak area distribution was randomly assigned to 100 singlets. These cumulative histograms quantify ^{2}D analyte peak widths, defined by the 95% inclusion after applying 2D mzCompare. (Left) In-phase with a 95% inclusion width of 15.7 ms. (Right) Out-of-phase with a 95% inclusion width of 19.0 ms.

Evaluation Against Statistical Overlap Theory



The 2D peak capacity and saturation factor will be enhanced by applying 2D mzCompare,

$$n_{c,2D\ mzCompare} = n_{c,2D}\left(\frac{^2w_b}{^2w_{b,cluster}}\right) \qquad \alpha_{2D\ mzCompare} = \alpha_{2D}\left(\frac{^2w_{b,cluster}}{^2w_b}\right)$$

where ^2w_b and $^2w_{b,cluster}$ represent an *average* ^{2}D peak width for the single component analyte peaks before and after applying 2D mzCompare, respectively

"global" estimate of the *enhancement ratio* given by $\frac{^2w_{b,cluster}}{^2w_b} \approx \frac{17ms}{200ms} \approx 0.085$

~12-fold enhancement

Saturation factor (α_{2D})		0.01	0.03	0.1	0.2	0.3	0.6	1.0
Singlet Components Found (%)	Average	100	98	95	91	87	76	62
	Std	2.2	2.0	2.2	2.6	3.5	4.2	4.0
Singlet Components Found and Identified (MV>800) (%)	Average	100	98	95	90	85	71	54
	Std	2.2	2.0	2.4	2.7	3.4	4.1	3.2

Conclusions and Future Work

An extended algorithm, 2D mzCompare designed for intra-chromatogram comparison to enable rigorous analyte discovery and identification, was developed for GC×GC-TOFMS data. Within the context of SOT, 2D mzCompare increases separation resolution by computationally minimizing the ^{2}D peak widths, enhancing 2D peak capacity and reducing the saturation factor. At low saturation factors (α_{2D}= 0.01, 0.03, and 0.1), over 95% of the simulated components are found to be mathematically resolved singlets (pure analyte components) by 2D mzCompare, while approximately 62% were found at α_{2D} = 1, exceeding predictions made by SOT. Using optimized 2D mzCompare parameters, improvements in ^2R_s and α_{2D} are about 12-fold, empirically validating SOT expectations. Furthermore, 2D mzCompare can be used as a preprocessing tool to determine or validate the "rank" (the number of analyte components) in overlapped regions of GC×GC-TOFMS chromatograms when combined with chemometric methods such as MCR-ALS or PARAFAC. Future studies will focus on applying the algorithm to real sample datasets.

References

Davis, J. M.; Carr, P. W. *Anal. Chem.* **2009**, *81*, 1198–1207.
Trinklein, T. J.; Synovec, R. E. *J. Chromatogr. A* **2022**, *1677*, 463321.
Cain, C. N.; Synovec, R. E. *Talanta* **2024**, *278*, 126453.
Ma, W.; Cain, C. N.; Synovec, R. E. *J. Chromatogr. A* **2026**, 466669.