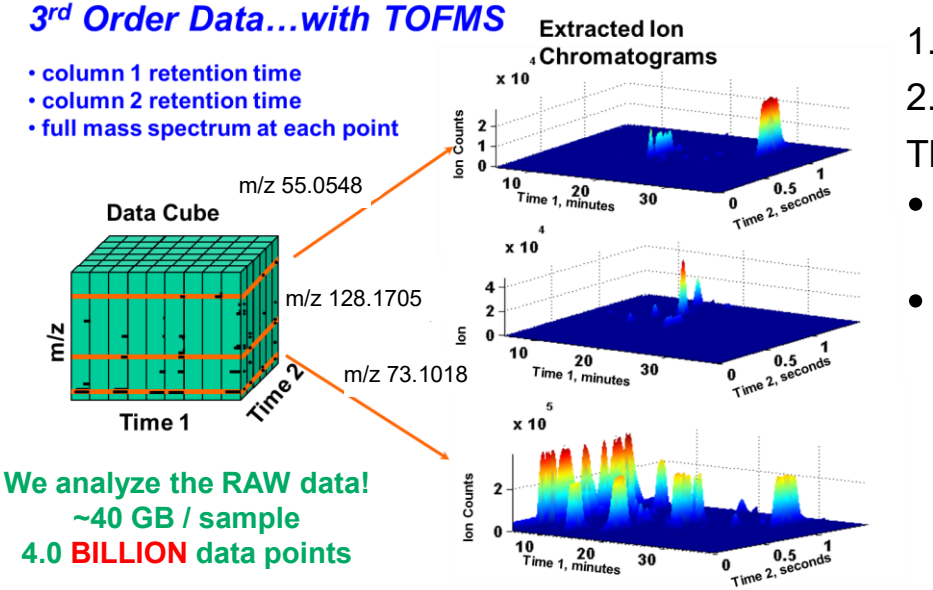
Discovery-based analysis for chromatographic trends using alteration analysis (ALA) and two-dimensional correlation analysis (2DCOR)

Chris E. Freye and Matthew J. Herman
High Explosives Science and Technology (Q-5)

1/13/2026
MDCW 2026

Approaching Complex Samples



Two major regimes of data analysis we want to avoid:

1. Visual comparison of chromatograms
2. Manual combination and analysis of peak tables

The alternative:

- Statistical methods for data reduction and/or feature selection
- Only spend time working up ***important*** analytes

Discovery-Based Chemometric Techniques

Classification

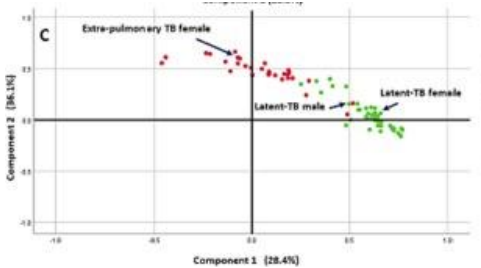
- PCA
- PLS-DA
- Random forest
- HCA

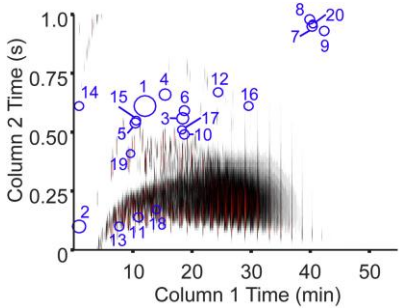
Feature Selection

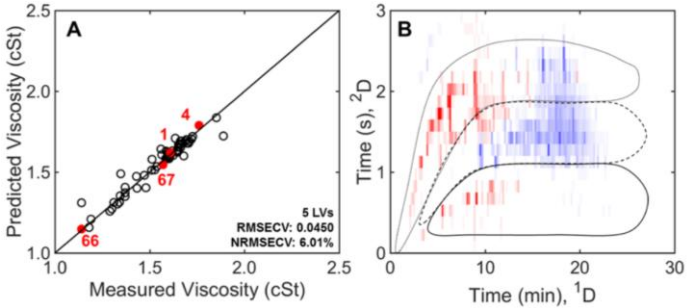
- Fisher ratio
- Pairwise (Fold-Change)
- Support vector machine (SVM)

Property Prediction

- PLS-R





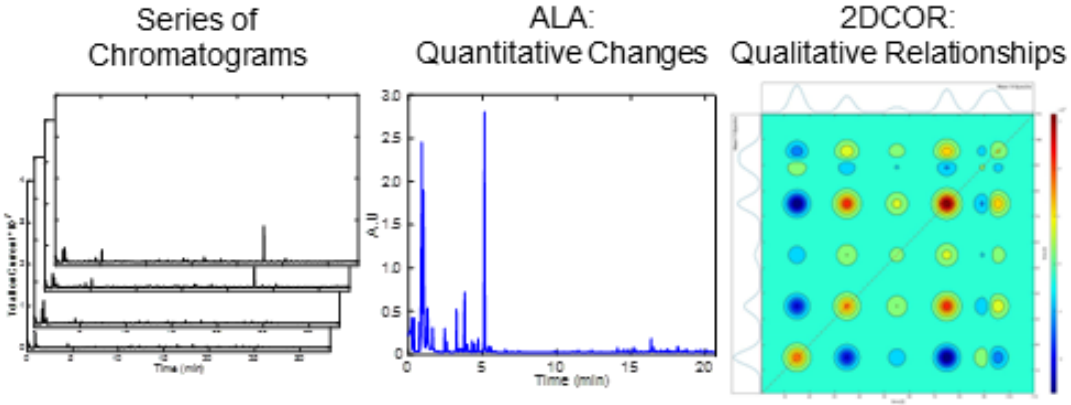


Current chemometric techniques are not suitable or sensitive enough to find trends across a series of samples

Chemometric Techniques for Uncovering Trends

Alteration Analysis (ALA) → 2D Correlation Analysis (2DCOR)

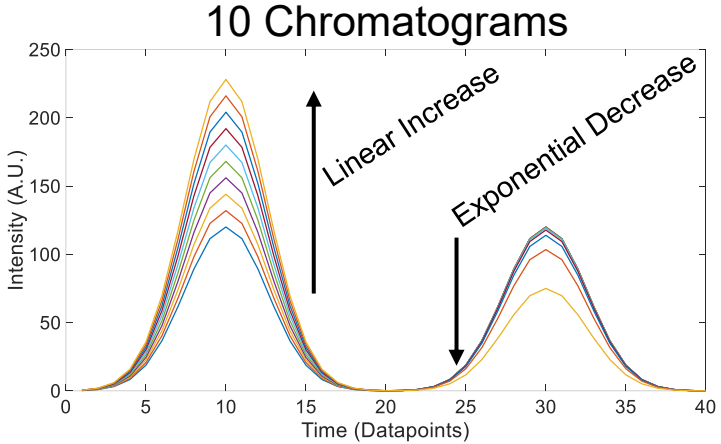
- Quantitative understanding of how individual datapoints vary across samples
  - Pinpoint chemical trends for a series of experiments
- Qualitatively determines relationships between data features and how these features change in respect to one another due to an applied external variable
  - Understand how chemical trends are related



ALA

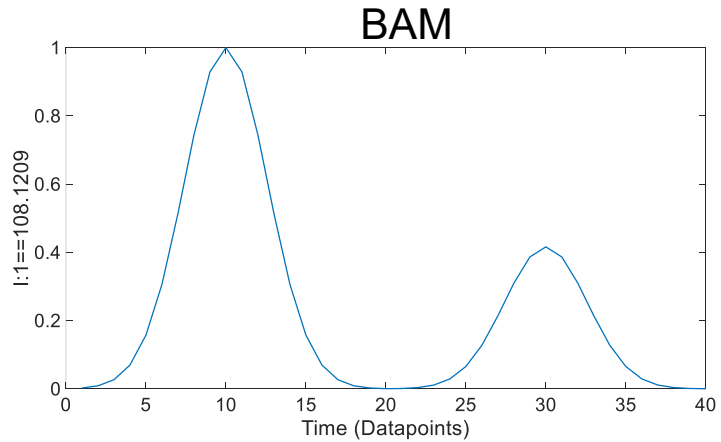
Series of data (chromatograms)

$$X = (x_{ij}) \qquad \begin{matrix} i = 1, 2, \dots, n \\ j = 1, 2, \dots, m \end{matrix}$$



Basic Alteration Map (BAM):
Overall change

$$b = (b_j) \qquad b_j = \max(x_j) - \min(x_j) \qquad j = 1, 2, \dots, m$$



ALA

Difference matrix

$$D = (d_{ij}) \qquad d_{ij} = x_{i+1j} - x_{ij} \qquad \begin{matrix} i = 1, 2, \dots, n-1 \\ j = 1, 2, \dots, m \end{matrix}$$

Synchronous Alteration Map (SAM): Linear change

$$s' = (s'_j) \qquad s'_j = \frac{b_j \overline{d}_j}{\sigma_{d_j} + 1} \qquad j = 1, 2, \dots, m$$

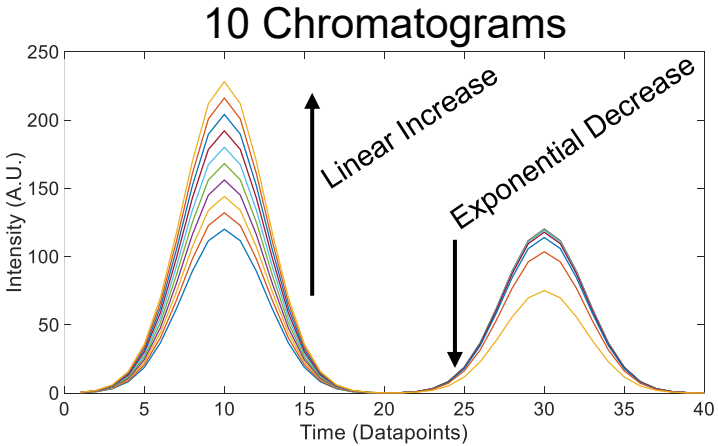
$$s = (s_j) \qquad s_j = \frac{s'_j}{\max(|s'|)} \qquad j = 1, 2, \dots, m$$

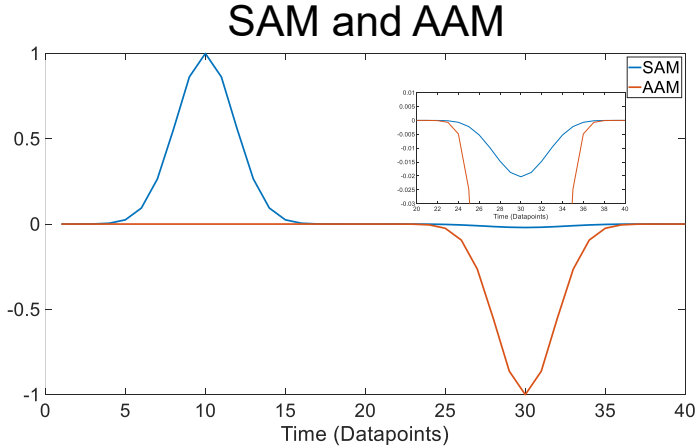
Asynchronous Alteration Map (AAM): Non-linear change

$$\boldsymbol{a}' = (a'_j) \qquad j = 1, 2, \dots, m$$

$$a'_j = (b_j - \left| \sum_{i=1}^{n-1} d_{ij} \right|) \sigma_{d_j} (\max(x_j) + \min(x_j) - 2\overline{x}_j)$$

$$\boldsymbol{a} = (a_j) \qquad a_j = \frac{a'_j}{\max(|a'|)} \qquad j = 1, 2, \dots, m$$

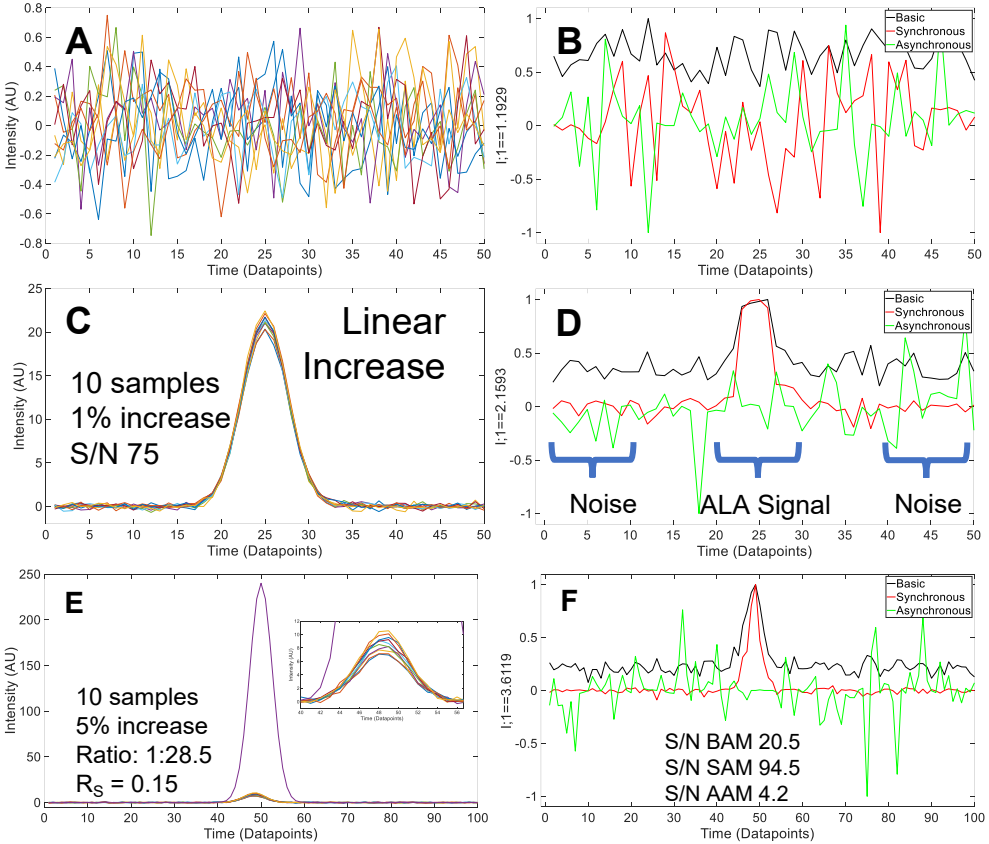




Los Alamos NATIONAL LABORATORY

Defining Success for ALA

- To evaluate quantitative performance of ALA, simulated noise, different amounts and types of change (linear, exponential, etc.)
- ALA was determined to be successful if a combination of BAM and one of the other two outputs from the ALA calculations had S/Ns greater than 10
- ALA works for peaks that are severely overlapped!



M.J. Herman, C.E. Freye, *Anal. Chem.* 2025, 97, 3, 1528–1538 https://doi.org/10.1021/acs.analchem.4c03660

2DCOR

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{I}(\boldsymbol{v},\boldsymbol{T}) - \bar{\boldsymbol{I}}(\boldsymbol{v}) & for \quad T_{min} \leq T \leq T_{max} \\ 0, & otherwise \end{cases}$$

Where **Y** is your final dynamic spectra matrix, $\bar{\boldsymbol{I}}(\boldsymbol{v})$ is the reference spectrum, and **T** is your external perturbation.

Synchronous correlation map (Φ (υ_1, υ_2))

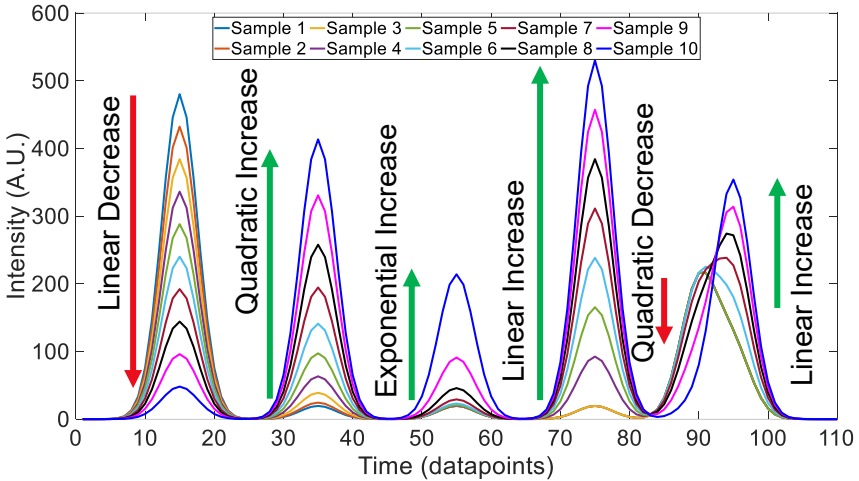
$$\Phi(\upsilon_1, \upsilon_2) = \frac{1}{n-1}\boldsymbol{Y}^T\boldsymbol{Y}$$

Asynchronous correlation map (ψ (υ_1, υ_2))

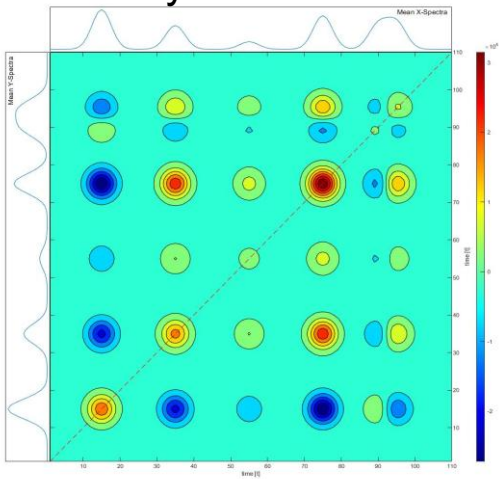
$$\psi(\upsilon_1, \upsilon_2) = \frac{1}{n-1}\boldsymbol{Y}^T\boldsymbol{N}\boldsymbol{Y}$$

$$N_{i,j} = \begin{cases} 0, & if\ i = j \\ \frac{1}{\pi}(j-i), & otherwise \end{cases}$$

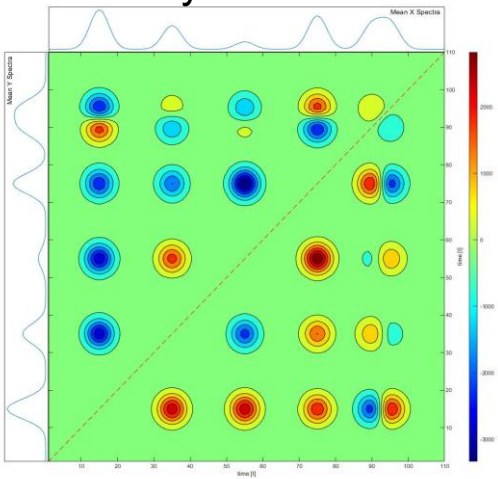
Application of 2DCOR



Synchronous



Asynchronous



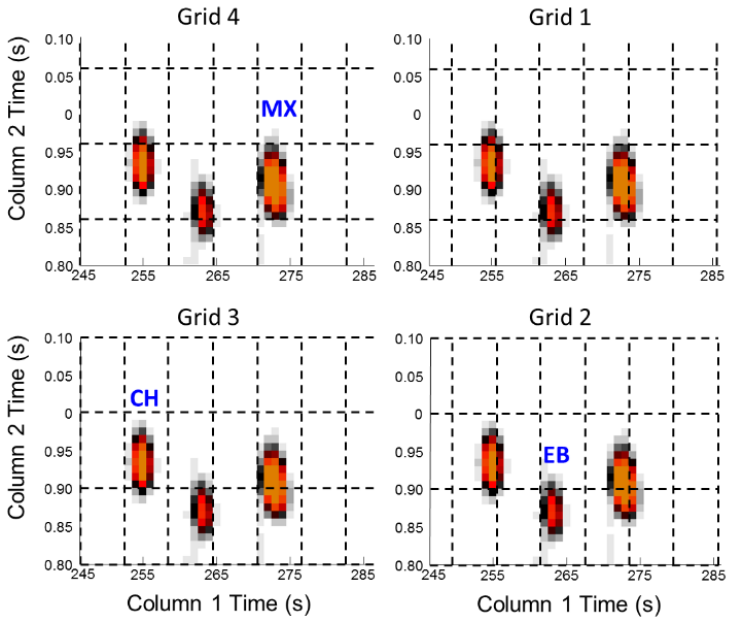
- Synchronous map provides information regarding the relative direction of coincidental change observed in the data set
- Asynchronous map characterizes the degree of coherence between signals measured at two different instances which are separated by a correlation time
- Using Noda's rule we can determine order of changes
  - 15 → 35 → 75 → 55 → 95 → 90

M.J. Herman, C.E. Freye, *Anal. Chem.* 2025, 97, 3, 1528–1538 https://doi.org/10.1021/acs.analchem.4c03660

Expansion of ALA and 2DCOR to GC×GC

- Chromatographic misalignment will reduce ability of ALA to discover chemical trends
  - Could be solved with alignment software, but this can be time consuming
- However, 2DCOR squares data size, even with pseudo-TIC chromatograms, the data size will be enormous (~70 billion datapoints)

Only Option is to Tile Dataset



Reproduced from B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Anal. Chem. 2015, 87, 7, 3812–3819

Previous ALA GC×GC Procedure

ALA Preprocessing and Computations

1. Normalization to internal standard
2. Tiling
3. Calculate ALA Parameters
4. Apply ALA S/N Threshold (S/N of 10)
5. Scale BAM, SAM, AAM to 1
6. Enforce requirement 2 of 3 ALA values (BAM, SAM, AAM)

Redundant Hit Removal

1. Pin hits by finding largest BAM within tile
2. Find top BAM, SAM, AAM m/z per tile
3. Remove redundant hits if within clustering window and share BAM, SAM or AAM m/z

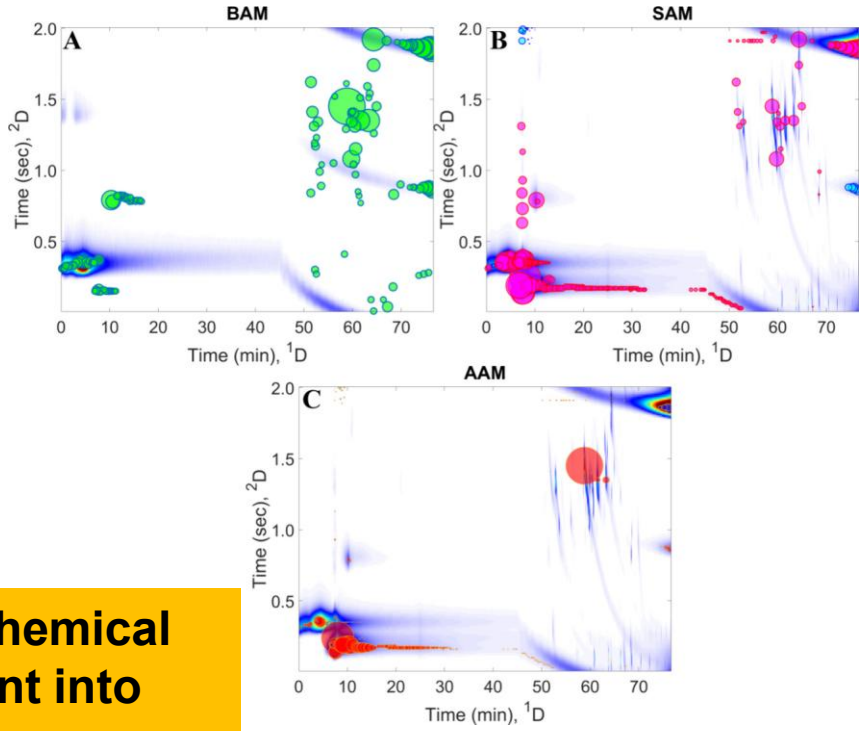
ALA Algorithm Output

- Hit list ranked by BAM values (or SAM or AAM)
- BAM, SAM, AAM Spectra
- Top m/z per hit

C.L. Ray and C.E. Freye, ACS Omega, 2025, 10, 26, 28404-28411

Recent Application of ALA to High Explosives

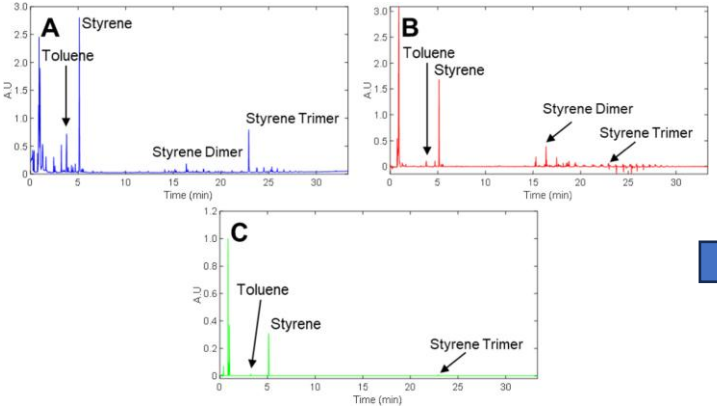
- Understand aging (chemical changes) of high explosives is critically important to our work
- F-ratio only partially worked since numerous intermediate decomposition products were formed
- Application of ALA discovered 250+ chemical changes but allowed insight into type of change

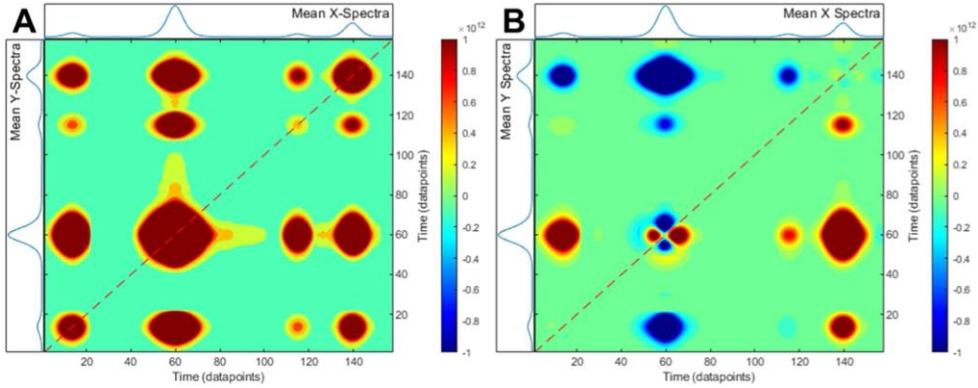


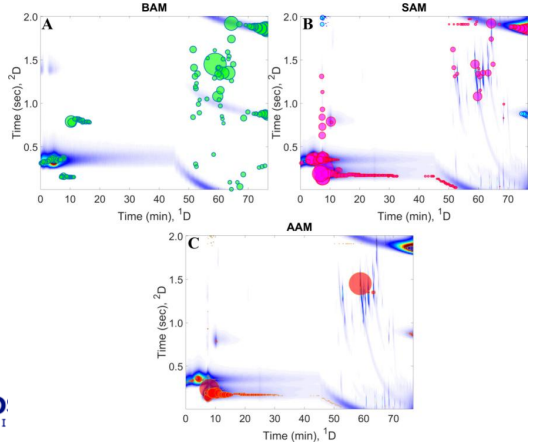
While ALA was essential to discovering chemical trends, a *significant* amount of effort went into understanding the relationship of the changes

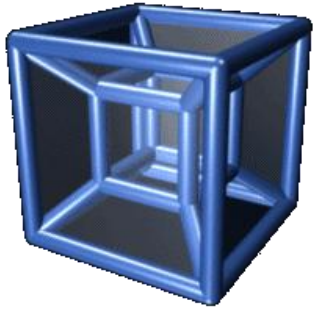
C.L. Ray and C.E. Freye, ACS Omega, 2025, 10, 26, 28404-28411

Problem with GC×GC and 2DCOR









Need to make our GC×GC data 1D

Refined ALA and 2DCOR Procedure for GC×GC

Initial ALA Algorithm Output

- Hit list ranked by BAM values (or SAM or AAM)
- BAM, SAM, AAM Spectra
- Top m/z per hit

Re-calculate ALA

- Re-center tile on pin location (peak maxima)
- Calculate BAM, SAM, AAM for new tile scheme
- Generate new hit list ranked by BAM values (or SAM or AAM)
- New BAM, SAM, AAM mass spectra
- Top m/z per hit

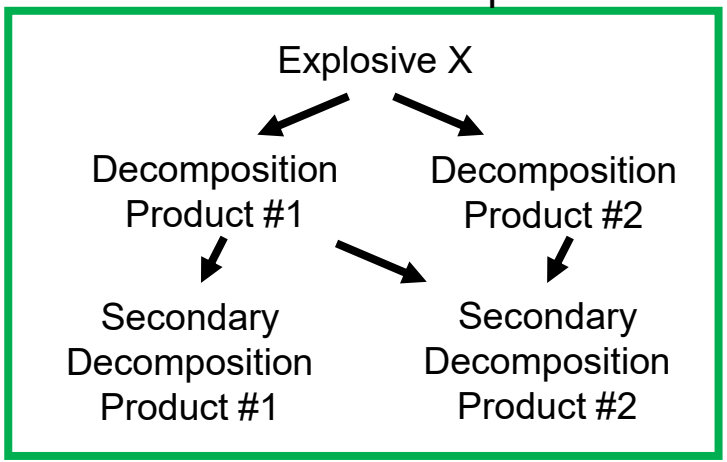
2DCOR

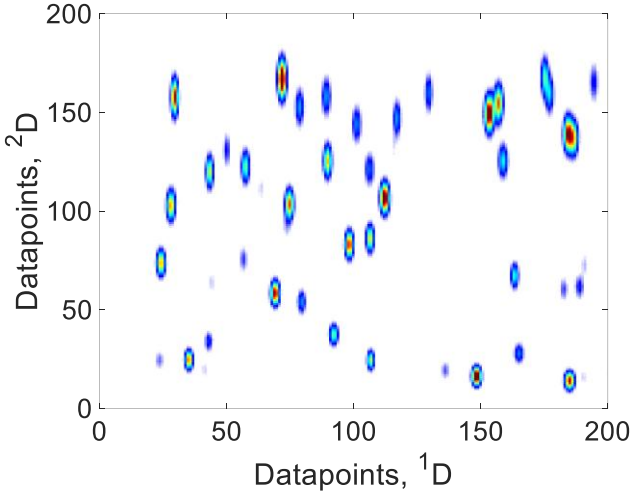
- Using raw data, sum intensities on both dimension for all m/z that have BAM value >0
- Calculate synchronous and asynchronous maps
- Determine order of changes

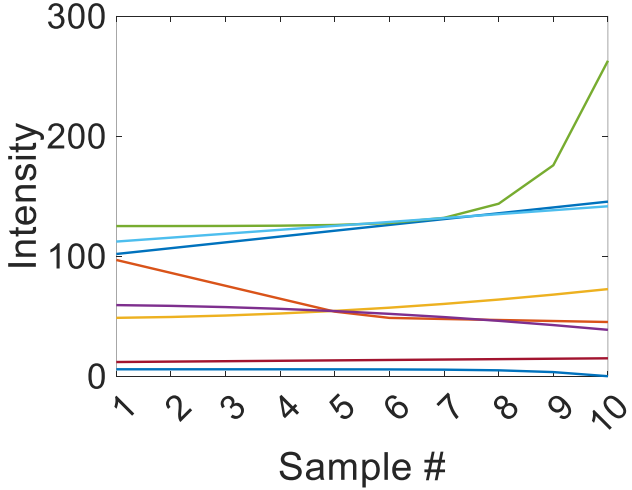
Let's see ALA and 2DCOR in Action

- Use in-silico dataset in order to simplify interpretation
  - Chromatogram with 50 peaks of which 8 peaks are changing at different rates
  - 10 chromatograms for the "experiment"

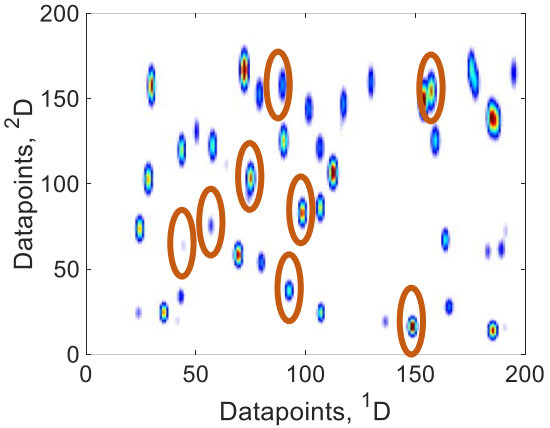
"Real-World" Example

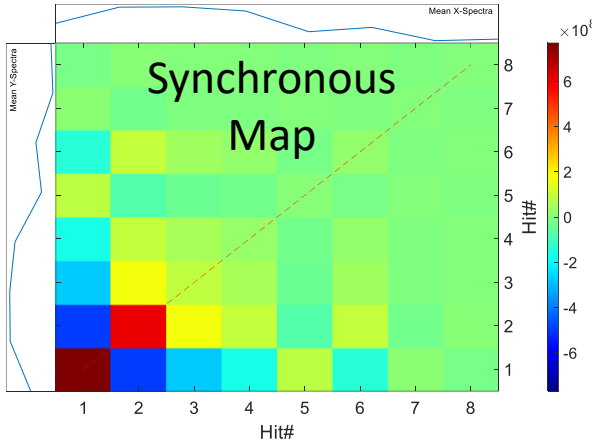


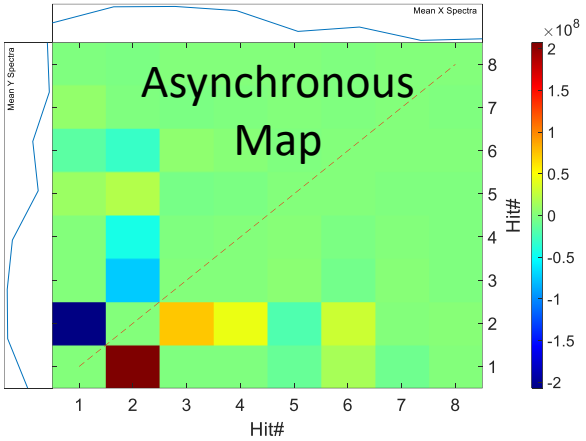




Let's see ALA and 2DCOR in Action





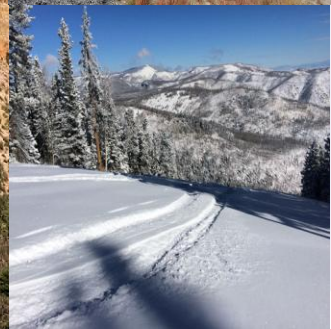


ALA Hit #	^2t_R	^1t_R	BAM	SAM	AAM	Top BAM m/z	Top SAM m/z	Top AAM m/z	Norm Sync Magnitude	Rank Order
1	155	157	1.0	0	0.003	113	113	221	1.0	7
2	17	148	0.82	0.06	1.0	187	187	88	0.79	2
3	104	75	0.37	1.0	0.001	7	7	55	0.12	6
4	83	98	0.24	0.42	0.001	12	12	30	0.04	5
5	152	79	0.21	0	0.001	136	3	136	0.01	8
6	38	92	0.16	0.05	0.001	247	247	248	0.03	1
7	62	44	0.05	0	0.001	245	1	245	0.001	3
8	75	57	0.03	0.005	0.001	58	58	15	0.001	4

ALA + 2DCOR informs of the significance of the change, how those changes are related, and the order of the changes

Acknowledgements

- ***Aging and Lifetimes Program***
  - *Mark Boggs*
- ***High Explosives and Energetics Program***
  - *Cameron Moore*
- ***LANL High Explosives Science and Technology (Q-5)***
  - *Colleen Ray and Michelle Corbally*



AGING AND LIFETIMES

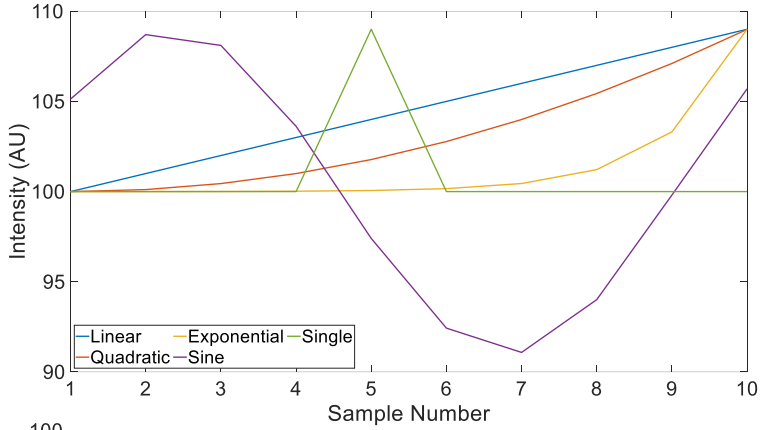
High Explosives and Energetics
No H.E., No Boom

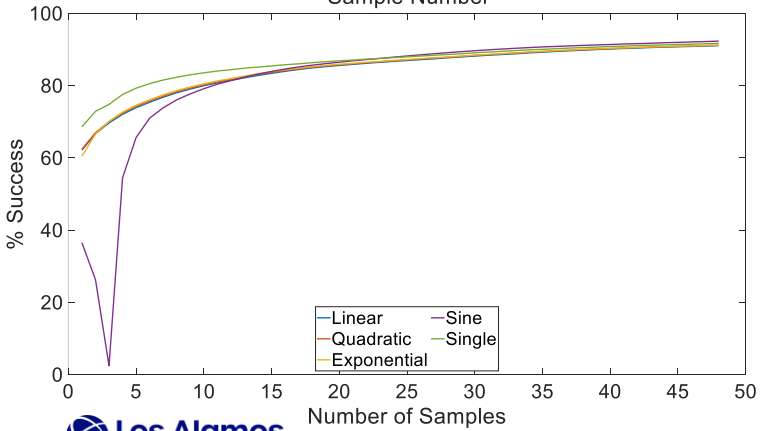
Los Alamos NATIONAL LABORATORY

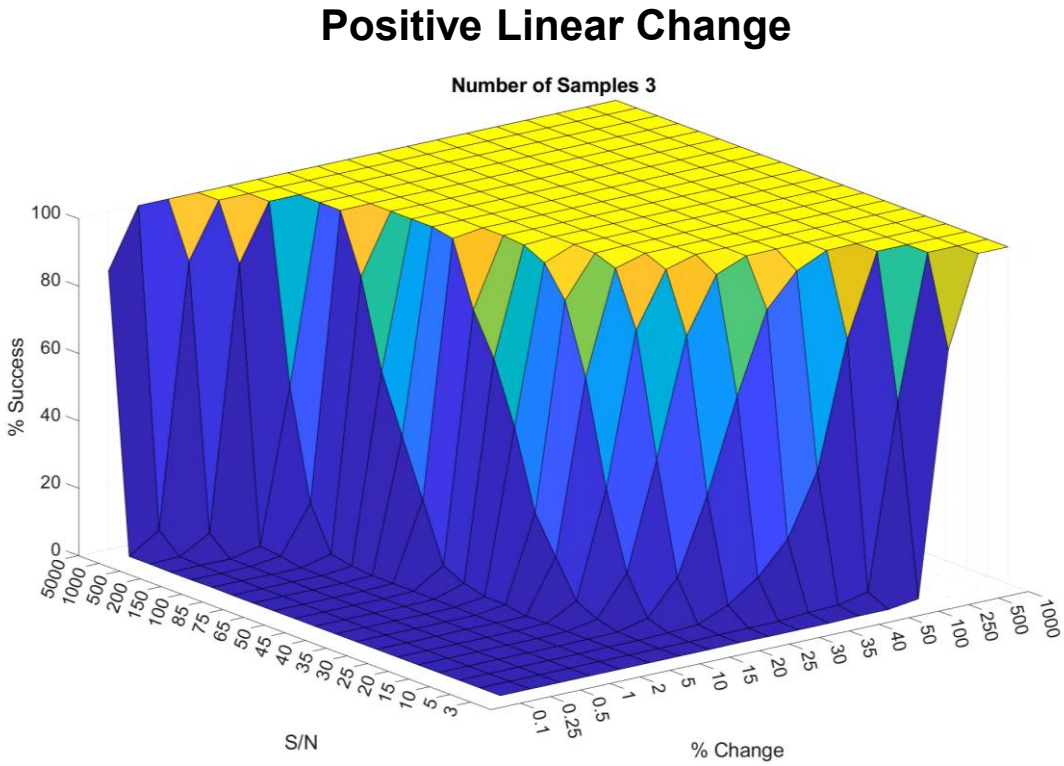
Questions?

M.J. Herman, C.E. Freye, “Expansion of Alteration Analysis and Two-Dimensional Correlation Analysis to Two-Dimensional Chromatographic Datasets” *in preparation*

Assessing ALA Success for Different Types of Changes

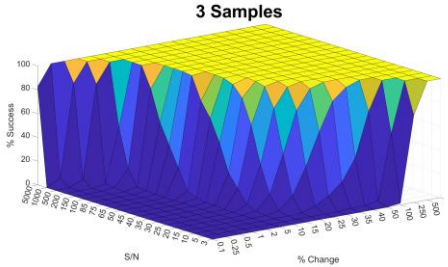


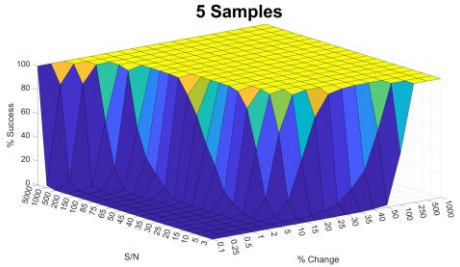


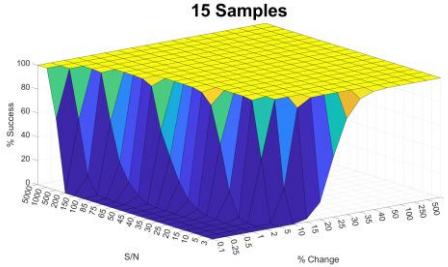


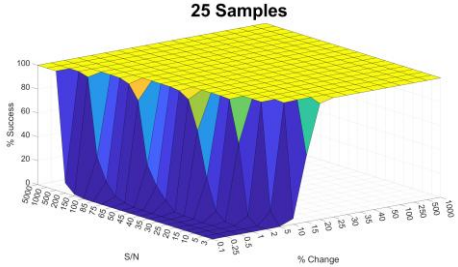
Los Alamos NATIONAL LABORATORY

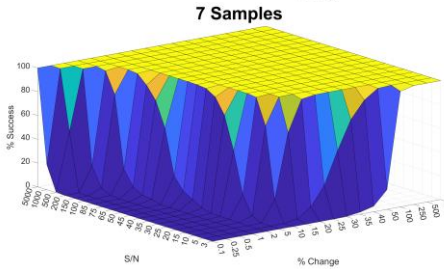
Linear Change (+)

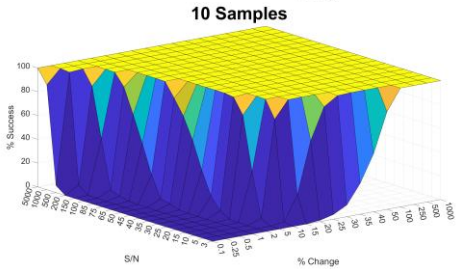


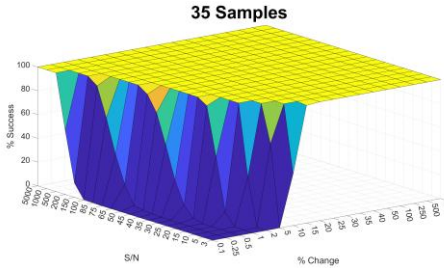


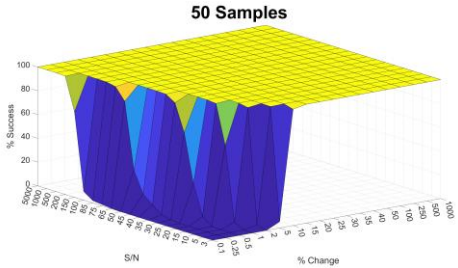






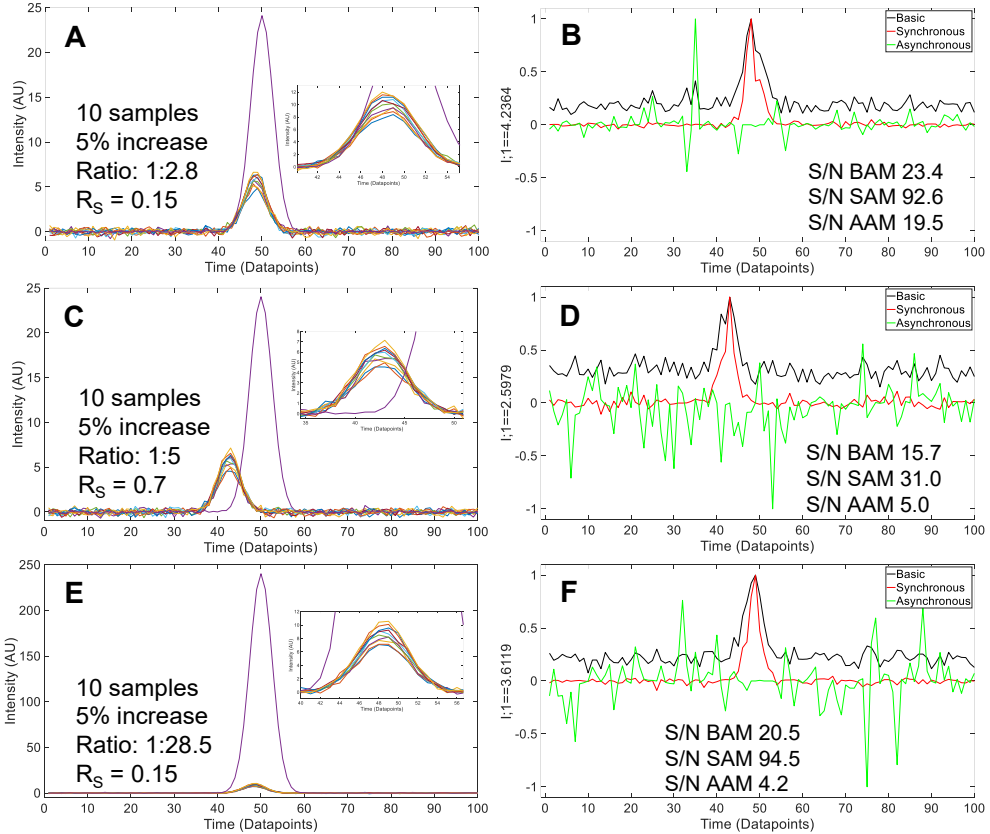






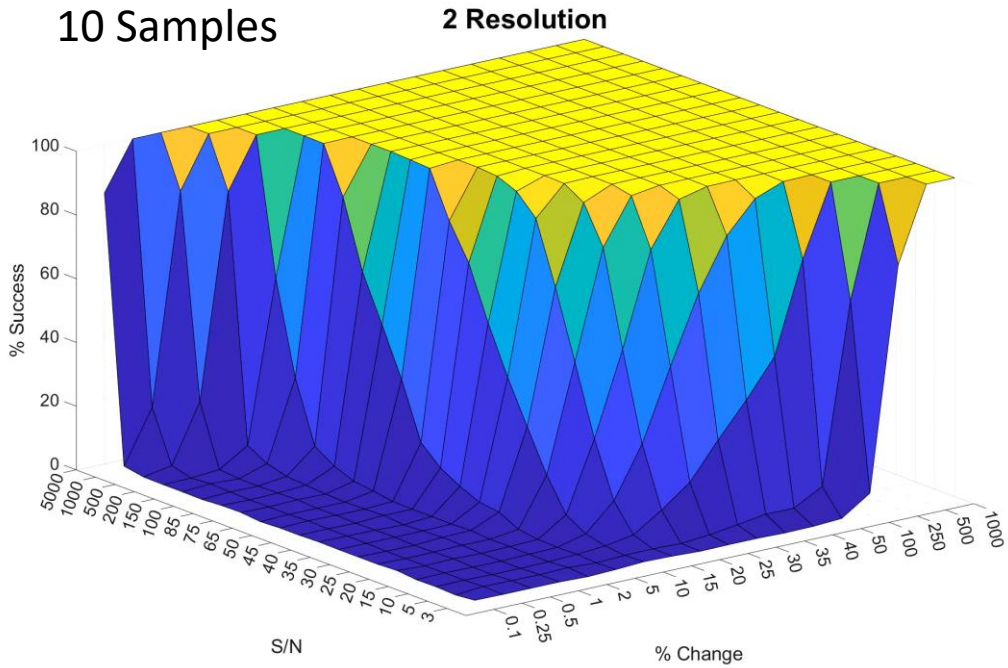
Evaluating ALA for Unresolved Peaks

- For ALA to be applicable to GC / GC×GC datasets it must be able to discover overlapped compounds
- As resolution increases and initial S/N increases, ALA probability of success increases
- Large interferents had less than expected impact on ALA success

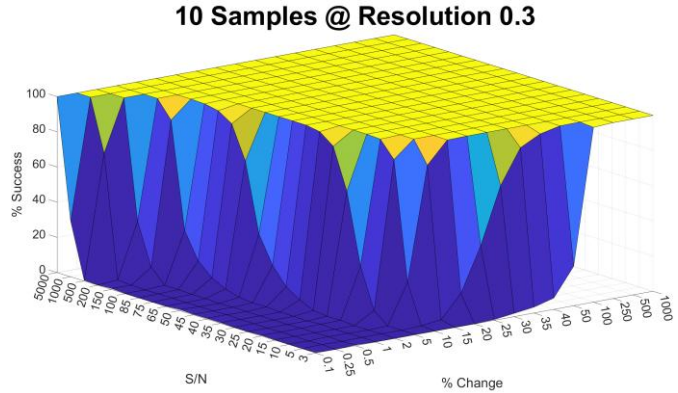


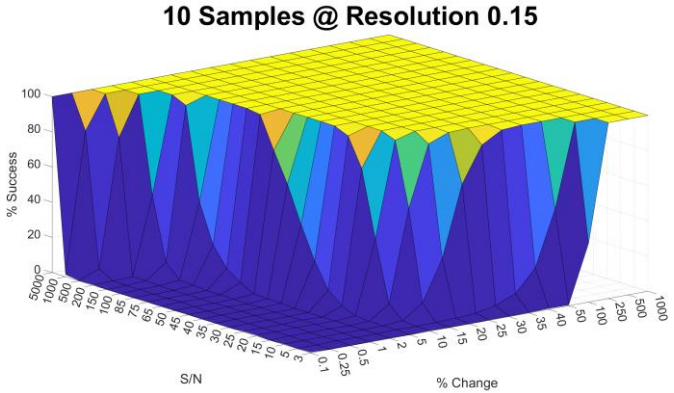
Evaluating ALA for Unresolved Peaks

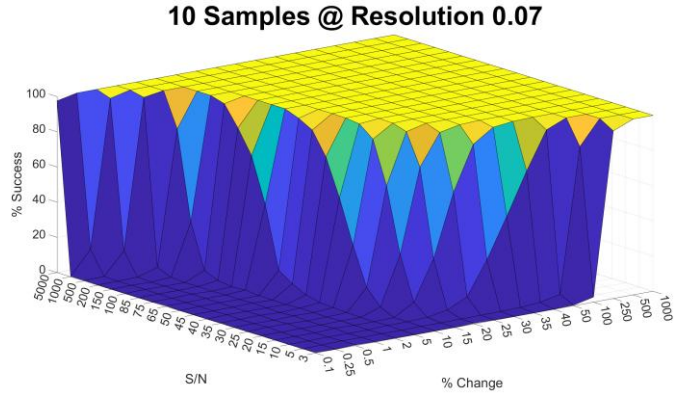
- ALA success is similar to a single peak down to a R_s of ~0.3
- ALA is successful all way down to R_s of 0.01
- As one would expect, as R_s decreases, the initial chromatographic S/N and the amount of change must increase

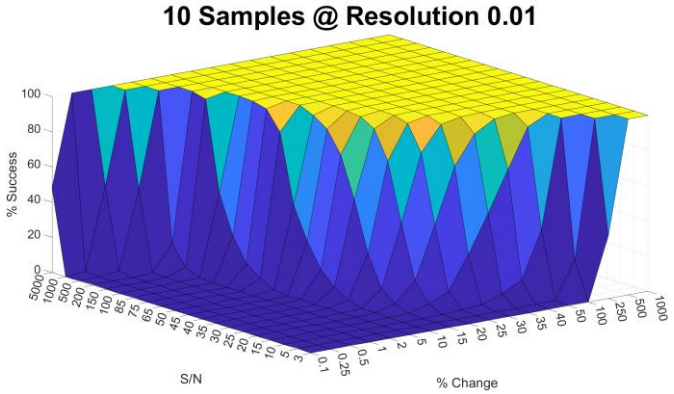


ALA Success at Different Resolutions for 10 Samples



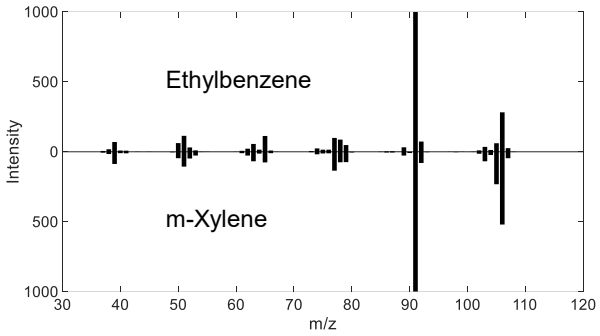


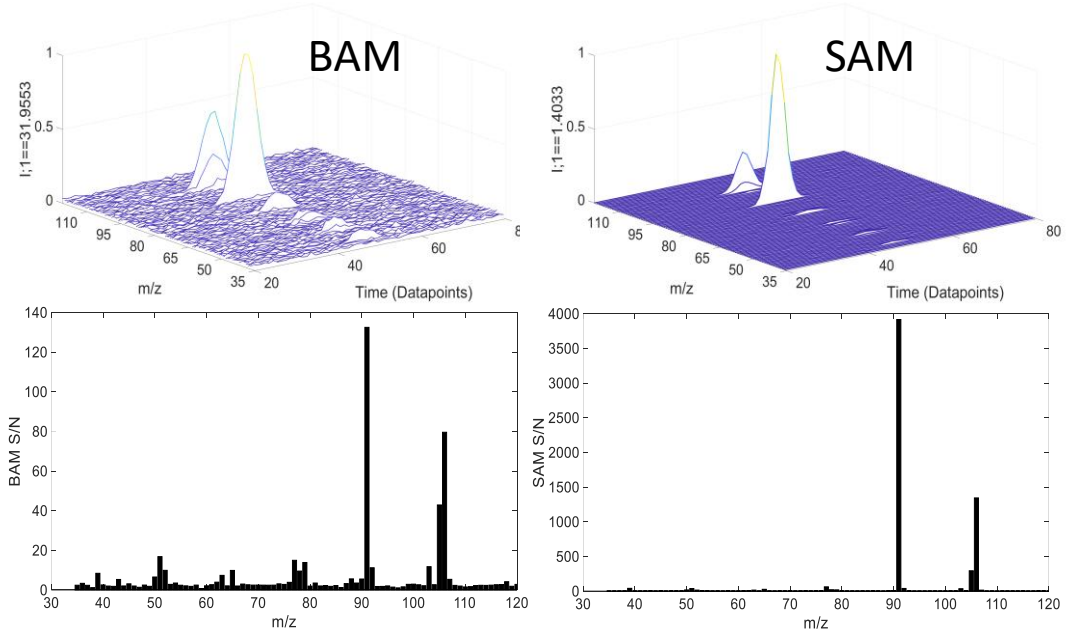




Expanding ALA to GC-TOFMS

- Evaluation of ALA on multivariate data was successful on highly overlapped, mass spectrally similar compounds
- Defining success if a single m/z passes the ALA S/N requirements

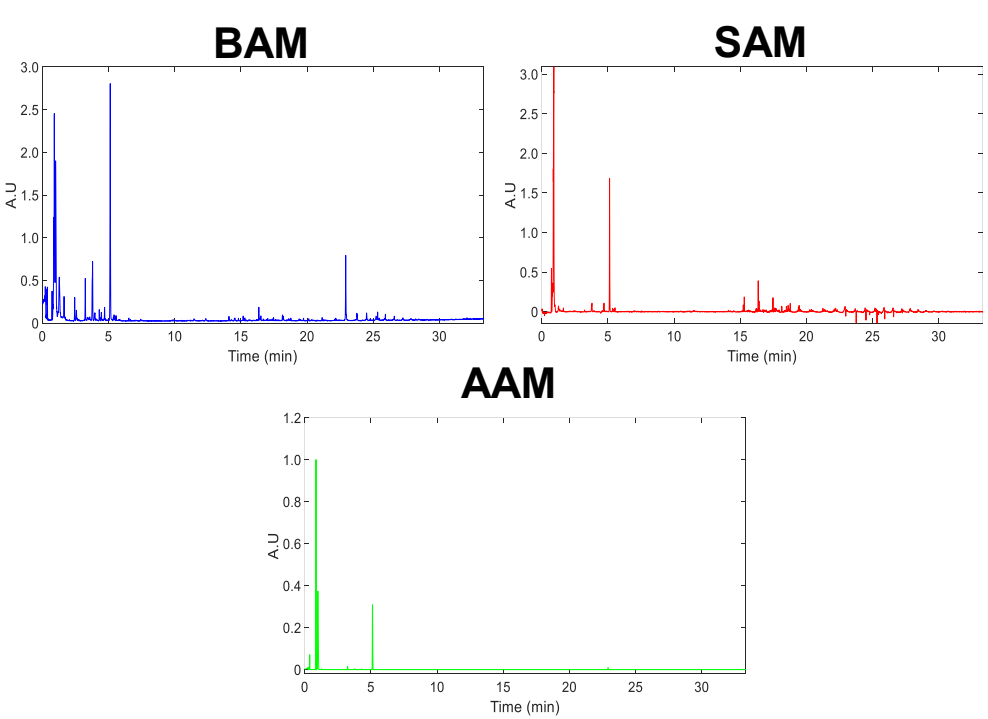


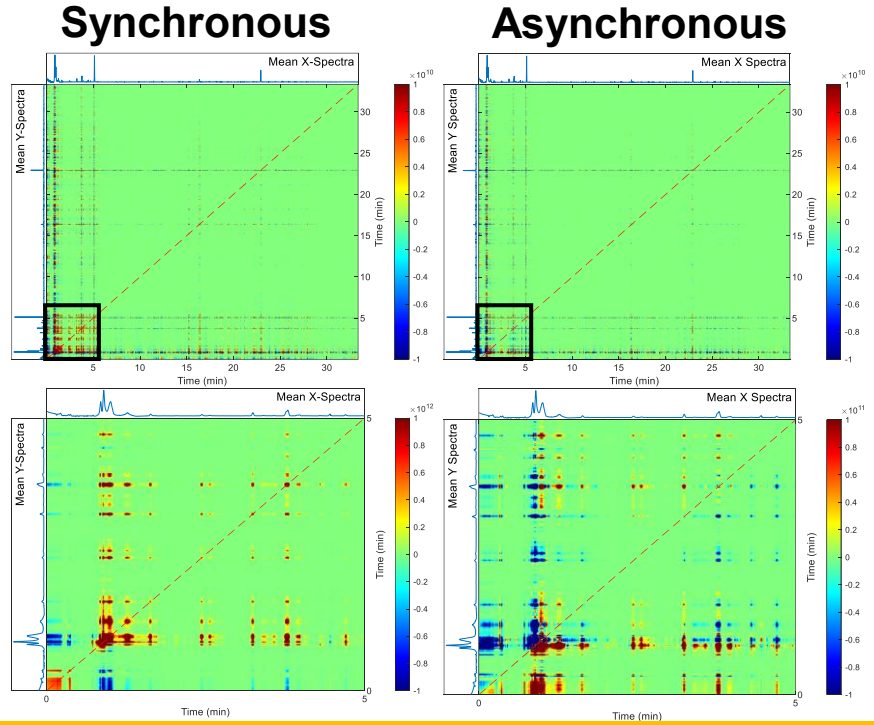


M.J. Herman, C.E. Freye, *Anal. Chem.* 2025, 97, 3, 1528–1538 https://doi.org/10.1021/acs.analchem.4c03660

8 m/z had S/N >10 for BAM and 15 m/z had S/N>10 for SAM

ALA and 2DCOR: Application to pyGC-TOFMS

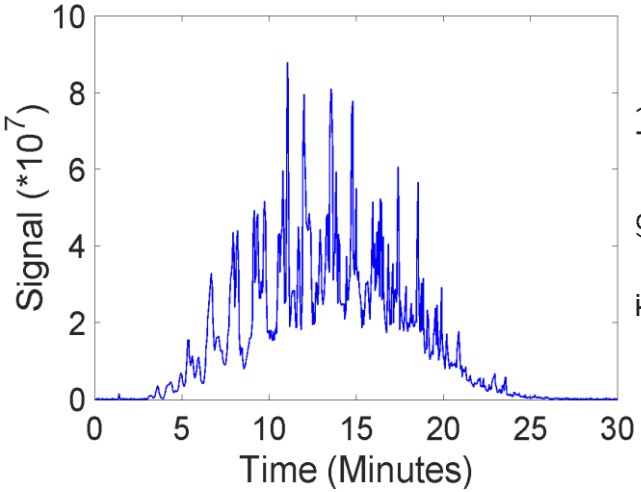




ALA and 2DCOR works on “real” GC-MS data

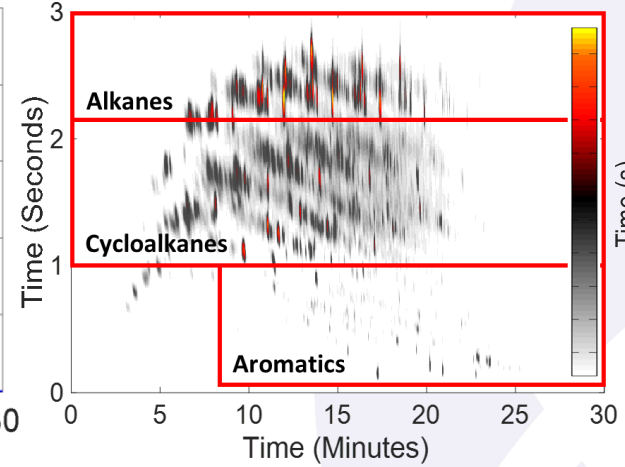
Disadvantages of 2-D Chromatography & High Resolution Mass Spectrometry

GC-TOFMS



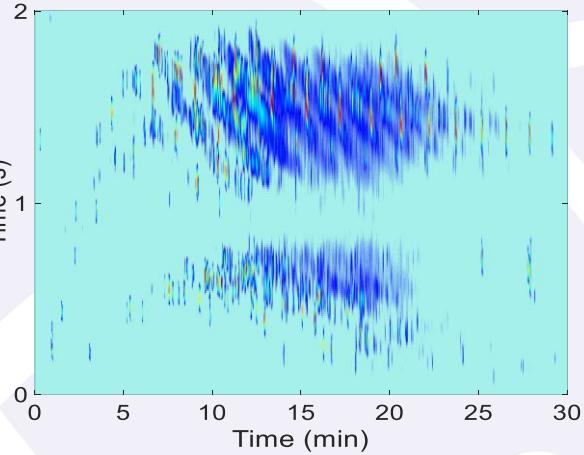
2,700,000 numbers

GC×GC-TOFMS



54,000,000 numbers

GC×GC-HRMS



4,000,000,000 numbers

GC-TOFMS datasets are ~54,000,000 datapoints
GC×GC-HRMS datasets are ~80,000,000,000 datapoints