Simple models for second-column retention-time variability across peaks from GCxGC

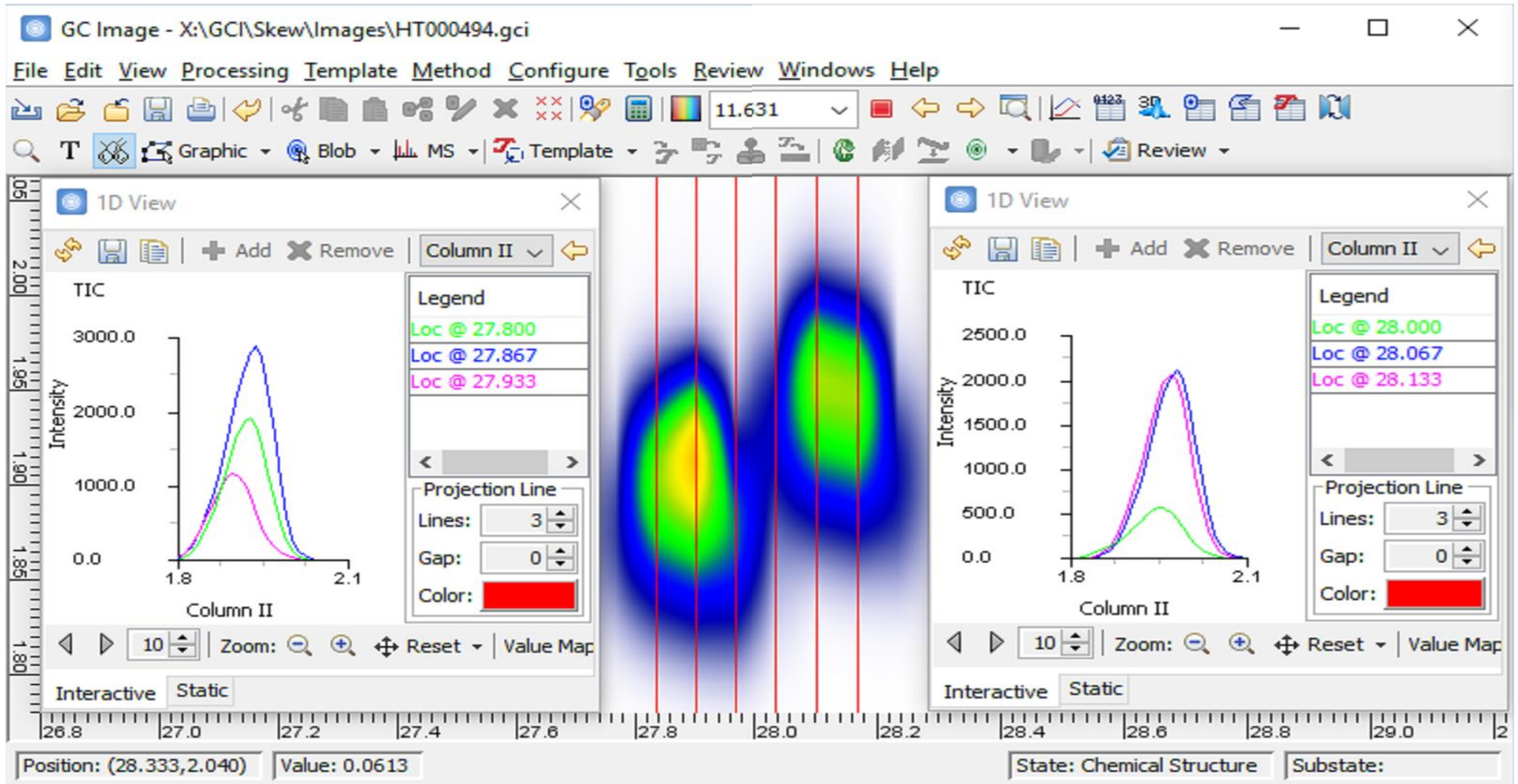
Stephen E. Reichenbach[1,2], **Davis W. Rempe**[1], Qingping Tao[2], and Chiara Cordero[3]

[1]University of Nebraska-Lincoln (Lincoln NE, USA)
[2]GC Image, LLC (Lincoln NE, USA)
[3]Università degli Studi di Torino (Torino, Italy)

8th Multidimensional Chromatography Workshop
Toronto ON, Canada, January 5-6, 2017

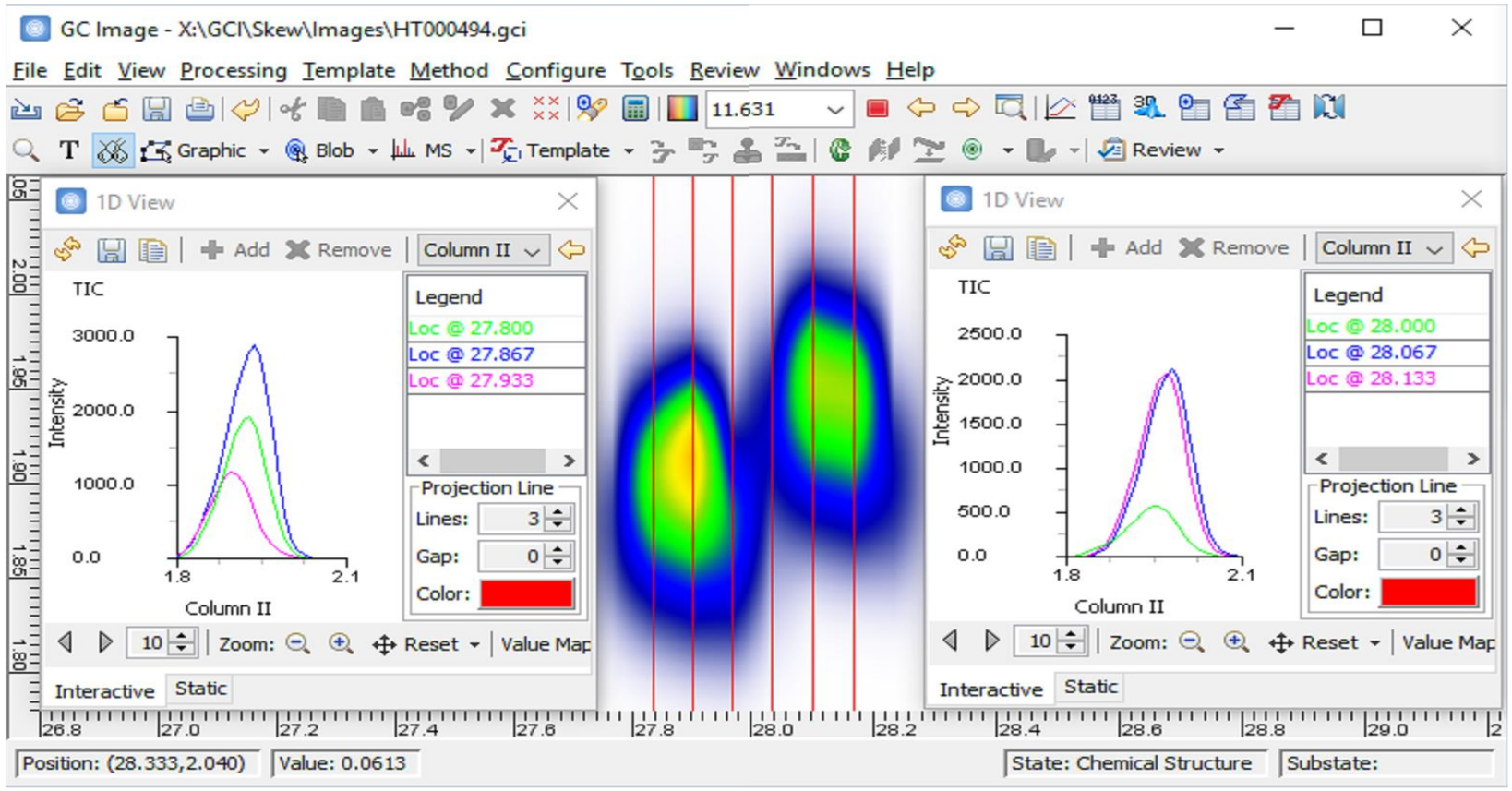
Background

- Second-dimension (^{2}D) separations of analytes create multiple 1D peaks across the first-dimension (^{1}D)
- ^{2}D chromatograms form columns of the 2D image
- ^{2}D peaks from each compound form 2D ***blobs***



Motivation

- Ideally, a compound's ^{2}D peaks identical ^{2}D retention times
- **In practice, that is not always the case**
- Shifting ^{2}D retention times pose problems for processing, e.g., blob detection & alignment



Motivation (2)

- **<u>Question</u>**: What factors relate to retention-time (RT) variability within a blob from one ^{2}D chromatogram to the next?
- Establish most important parameters to predict RT shift through simple models that are:
  - Easily implemented
  - Robust

Outline

- Background
- Motivation
- Analysis
- Results
- Conclusion

Outline

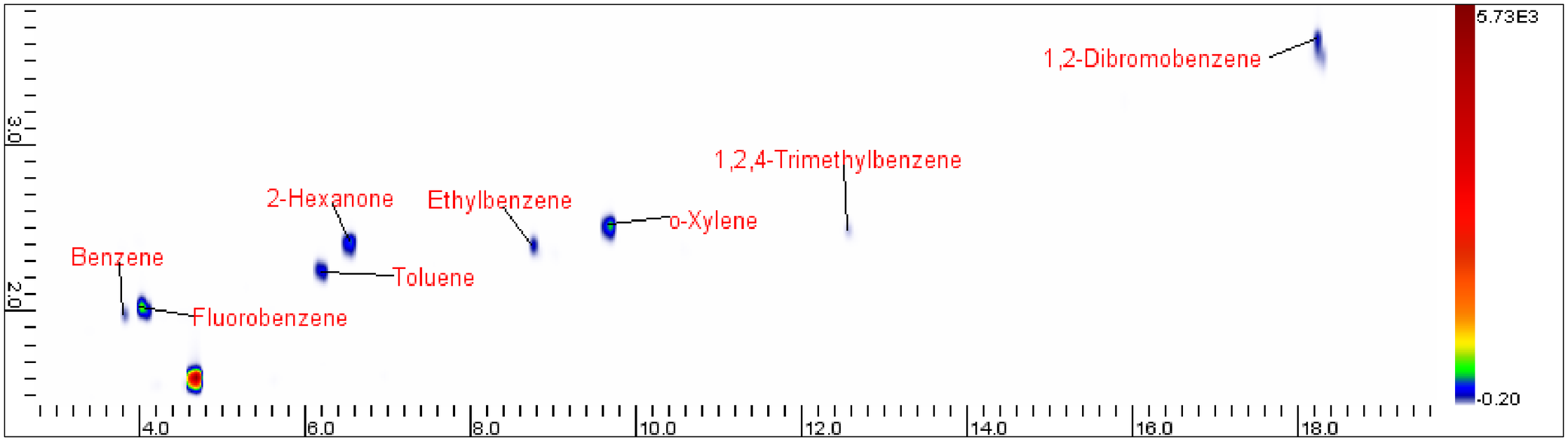
- Background
- Motivation
- Analysis
- Results
- Conclusion

Data Analysis Overview

- For chromatograms of interest:
  - Measure ^{2}D RT shifts within analyte blobs
  - Assess relationships between ^{2}D shifts and ^{2}D peak attributes
  - Build models based on relationships:
    - Univariate, bivariate
    - Linear, quadratic
  - How well do models predict shift?
- 4 different GCxGC datasets
  - Calibration
  - Chemical-Agents
  - Temperature-Ramp
  - Gas-Pressure

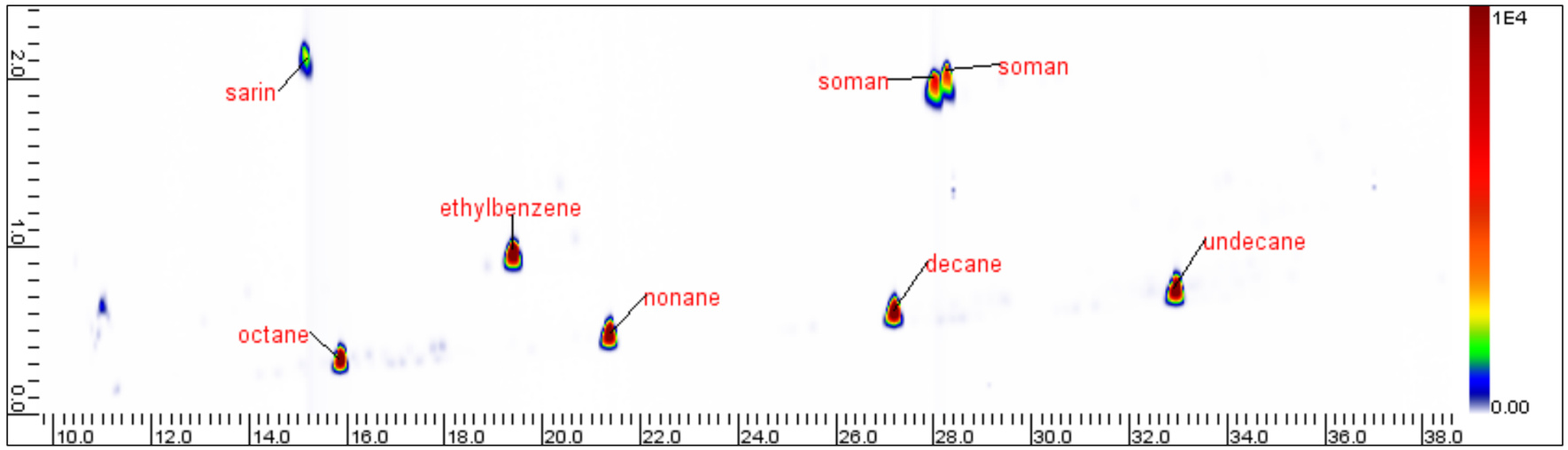
Datasets: Calibration

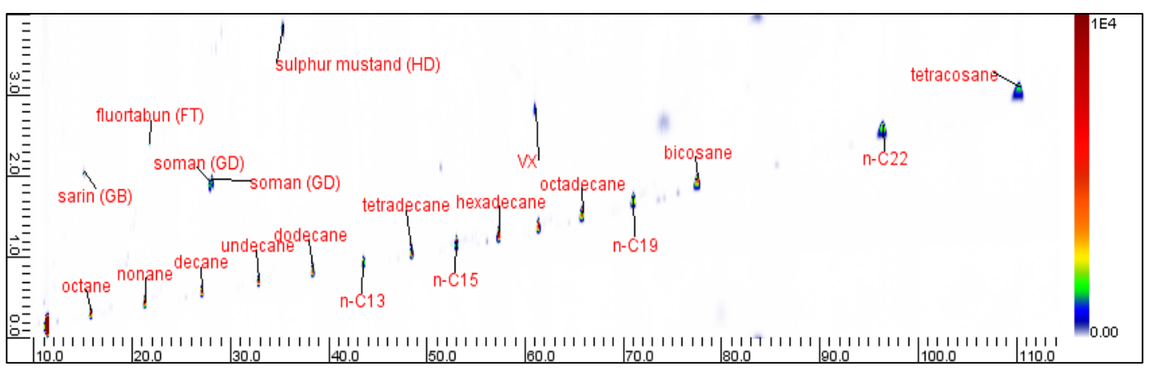
- 25 GCxGC chromatograms with varying concentrations
  - Five replicate runs for each of five calibration vials for ASTM D5580-02, “Standard Test Method for Determination of Benzene, Toluene, Gasoline by Gas Chromatography”

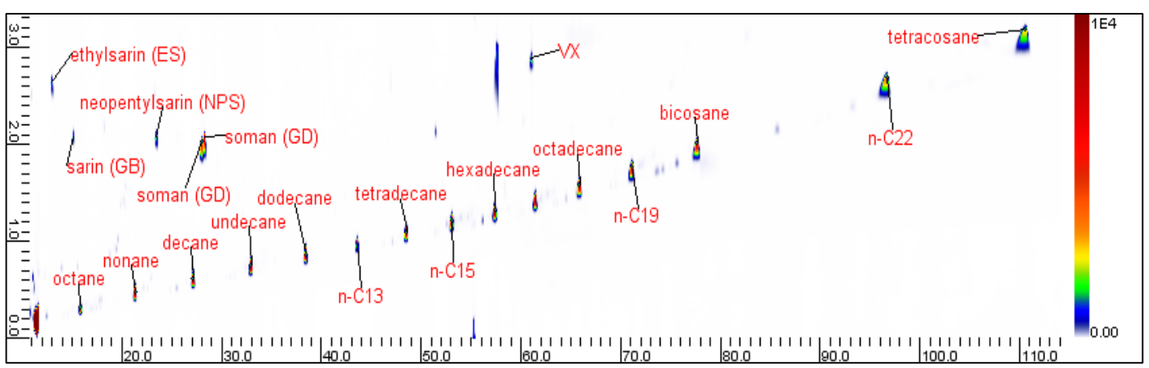


Datasets: Chemical-Agents

- 3 GCxGC chromatograms, each with different composition

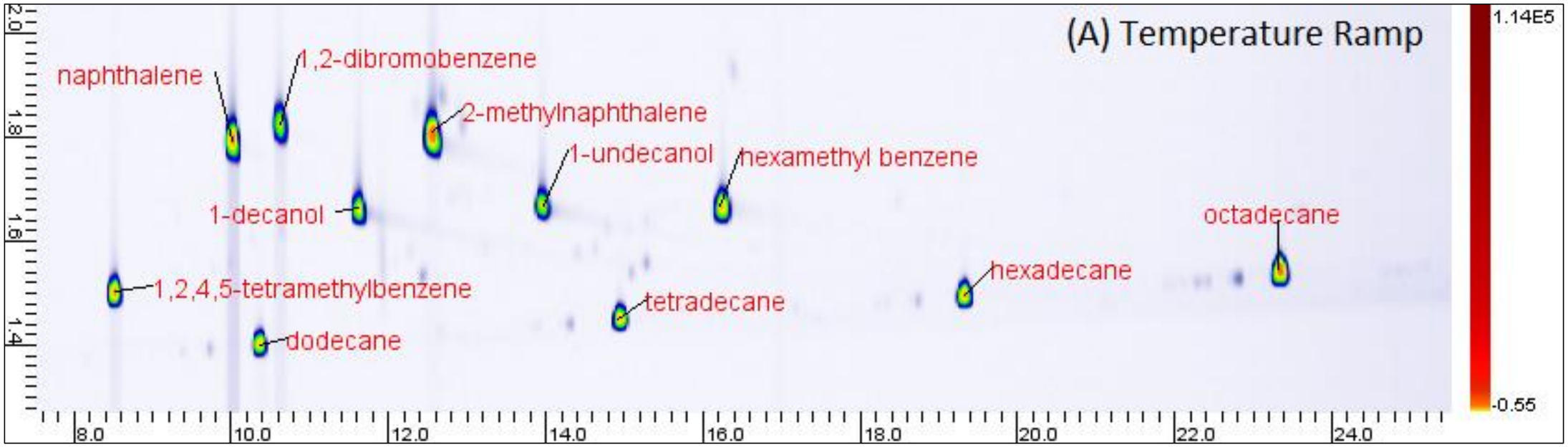






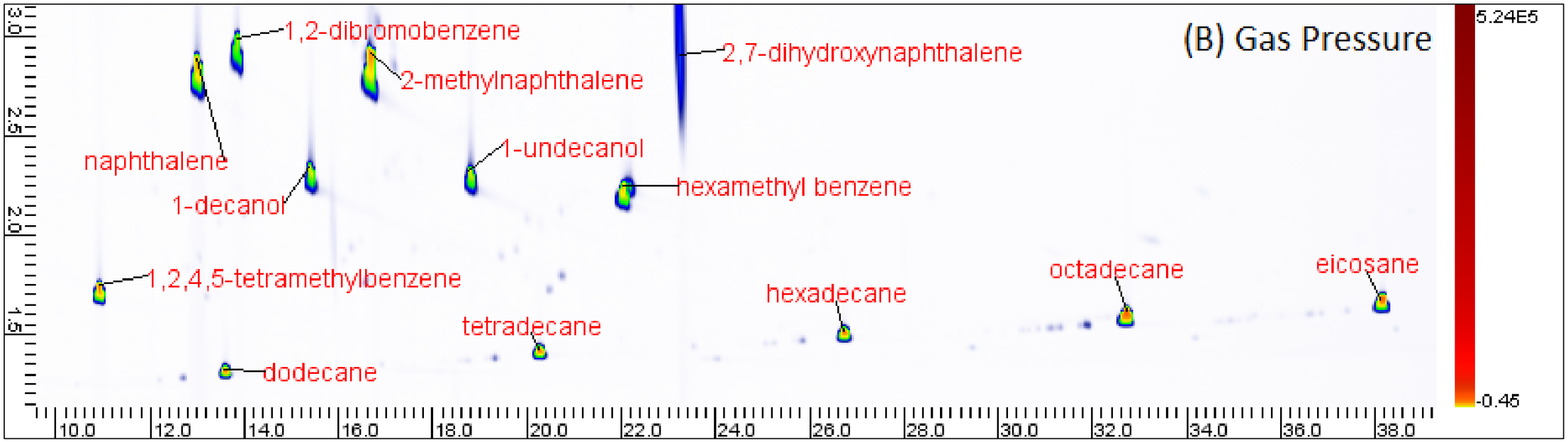
Datasets: Temperature-Ramp

- 10 GCxGC chromatograms, each with different temperature-ramp rates (from 2 to 11 °C/min)



Datasets: Gas-Pressure

- 8 GCxGC chromatograms, each with different inlet gas pressure (from 17 to 24 psi)



Analysis Metric

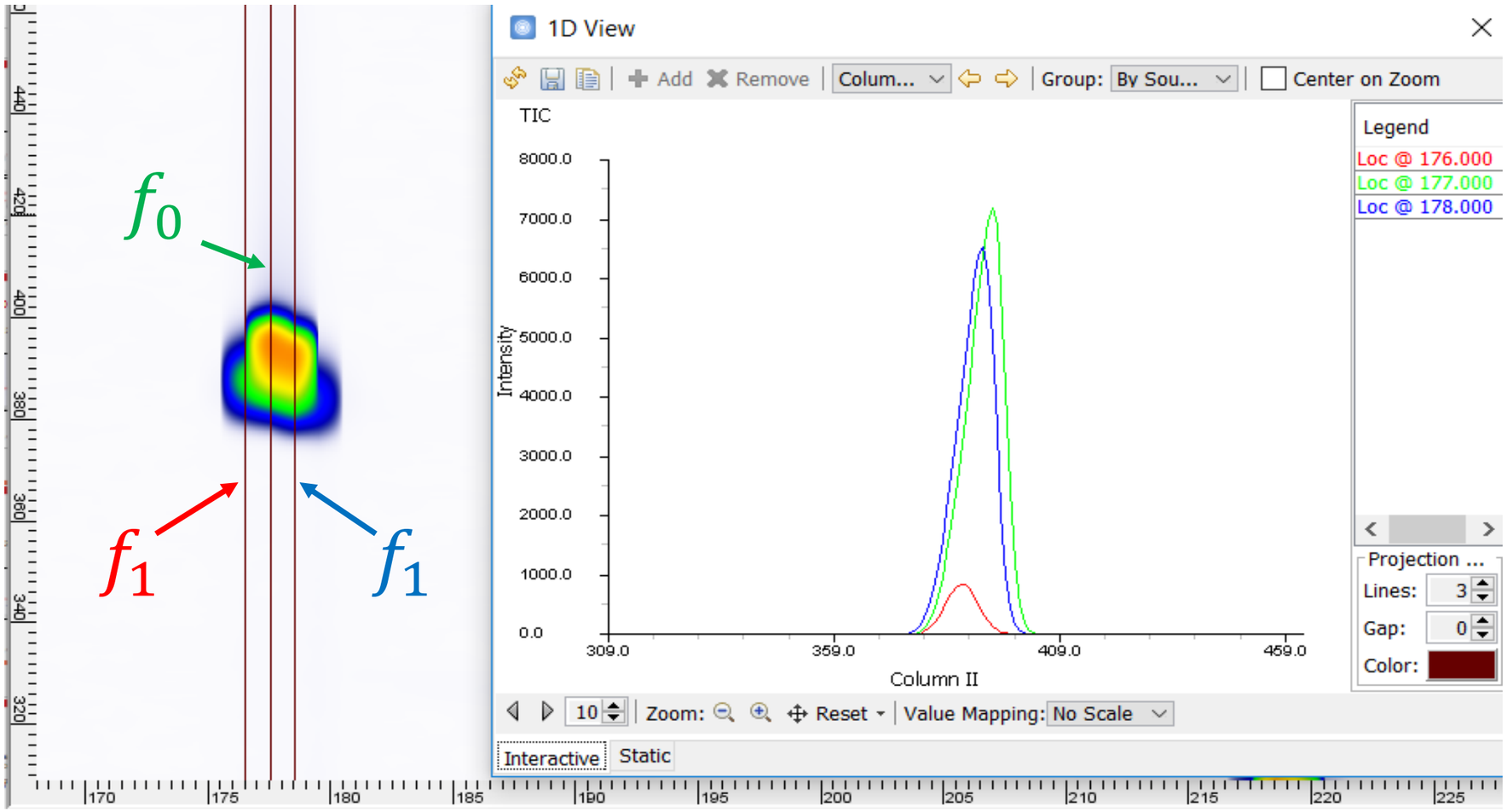
- RT difference is the dependent variable
- Relationship to various independent variables
- RT difference for consecutive ^{2}D peaks, f_0 and f_1, is relative shift that maximizes cross-correlation:

$$\Delta t = \max_s \sum_{i=-n}^{n} \left(\bar{f}_0(i) \cdot \bar{f}_1(i+s) \right)$$

- $\bar{f}(i)$ is standardized detector response (intensity)
- $i = 0$ is apex, and $n = 2 \cdot {}^2\sigma$
- Δt values shown here are normalized by blob's ${}^2\sigma$

Analysis Metric (2)

- Compute Δt for
  - Increasing: ^{2}D peak previous to apex and ^{2}D apex peak (f_1 & f_0)
  - Decreasing: ^{2}D apex peak and next ^{2}D peak following apex (f_0 & f_1)
- f_0 is always the apex peak



Factors Considered

- Primary independent variables
  - Retention time at apex (${}^{1}t_r$ and ${}^{2}t_r$)
  - Total intensity of each peak (${}^{2}\eta_0$ and ${}^{2}\eta_1$)
  - Standard-deviation width of each peak (${}^{2}\sigma_0$ and ${}^{2}\sigma_1$)
  - Skewness of each peak (${}^{2}\gamma_0$ and ${}^{2}\gamma_1$)
- Secondary variables from combining primary variables
  - Difference of total intensities (${}^{2}\eta_0 - {}^{2}\eta_1$)
  - Difference of standard deviation widths (${}^{2}\sigma_0 - {}^{2}\sigma_1$)
  - Difference of skewness values (${}^{2}\gamma_0 - {}^{2}\gamma_1$)

Simple Models

- Fit models for RT difference vs. various factors
  - Linear, low-order polynomial, both univariate, bivariate, e.g.:
    - $\Delta t_{est} = a \times {}^{2}\eta_0 + b$ [Linear, univariate]
    - $\Delta t_{est} = a \times \left({}^{2}\eta_0\right)^2 + b \times {}^{2}\eta_0 + c$ [Quadratic, univariate]
    - $\Delta t_{est} = a \times {}^{2}\eta_0 + b \times {}^{2}\eta_1 + c$ [Linear, bivariate]
- Predict RT difference between two peaks
- Variability across dataset is root-mean-square (RMS) of RT differences of multiple peaks
- Measure reduction in RMS for each model

Outline

- Background
- Motivation
- Analysis
- Results
- Conclusion

Constant Model (from mean RT shift)

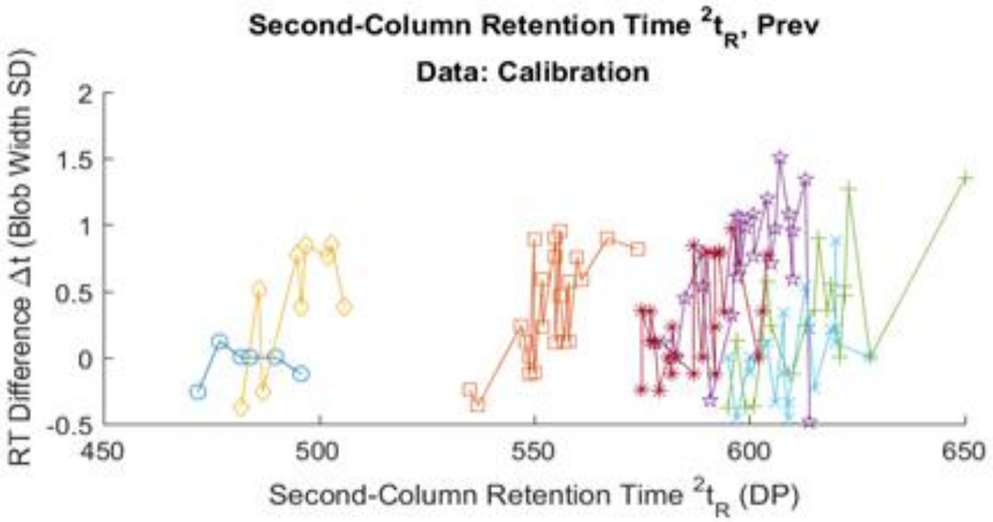
- Mean RT shift of peak-pairs builds “constant” model
  - $\Delta t_{est} = \Delta t_{avg}$
- Small RMS reduction using all pairs of ^{2}D peaks
- Improvement using separate constant models for:
  - Increasing: previous $\rightarrow$ apex
  - Decreasing: apex $\rightarrow$ next

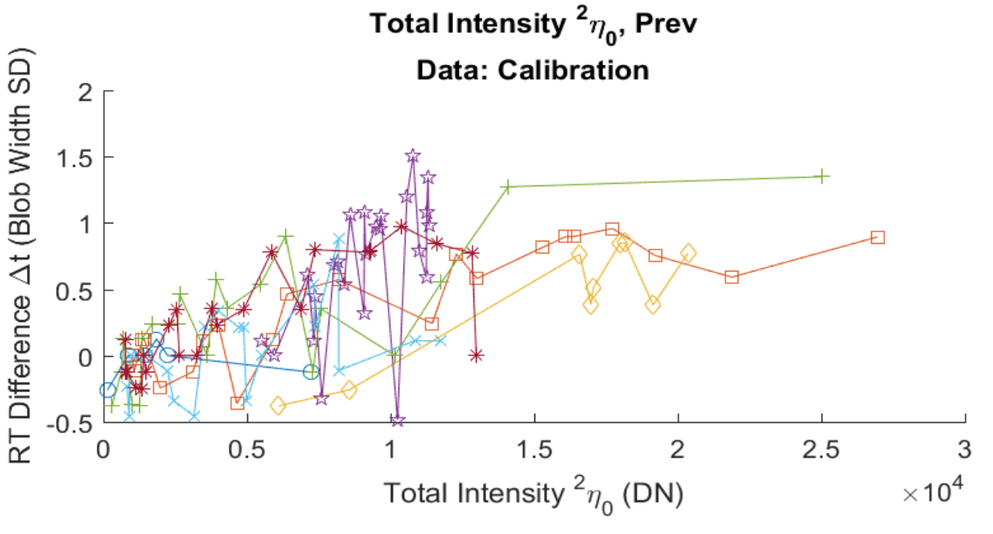
Dataset	All Pairs			Increasing Intensity			Decreasing Intensity		
	RMS	Mean	St.Dev.	RMS	Mean	St.Dev.	RMS	Mean	St.Dev.
Calibration	0.92	−0.39	0.83	0.57	0.33	0.47	1.12	−0.99	0.53
Chem.Agts.	0.45	−0.04	0.45	0.38	0.30	0.24	0.51	−0.38	0.34
Temp.Ramp	0.51	−0.07	0.50	0.44	0.32	0.31	0.57	−0.46	0.33
Gas-Press.	1.01	−0.10	1.01	0.84	0.72	0.43	1.17	−1.04	0.55

Independent Variables

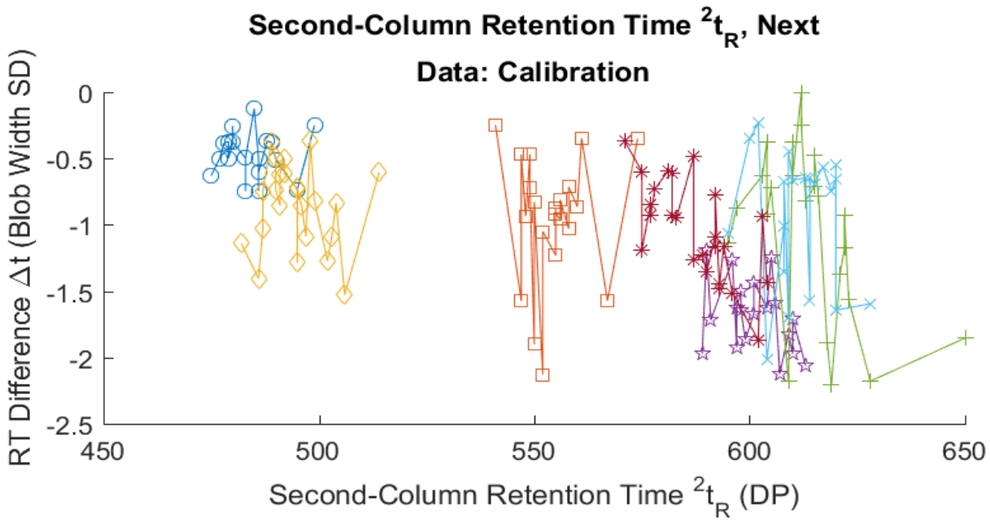
- Investigated relationships for all primary variables

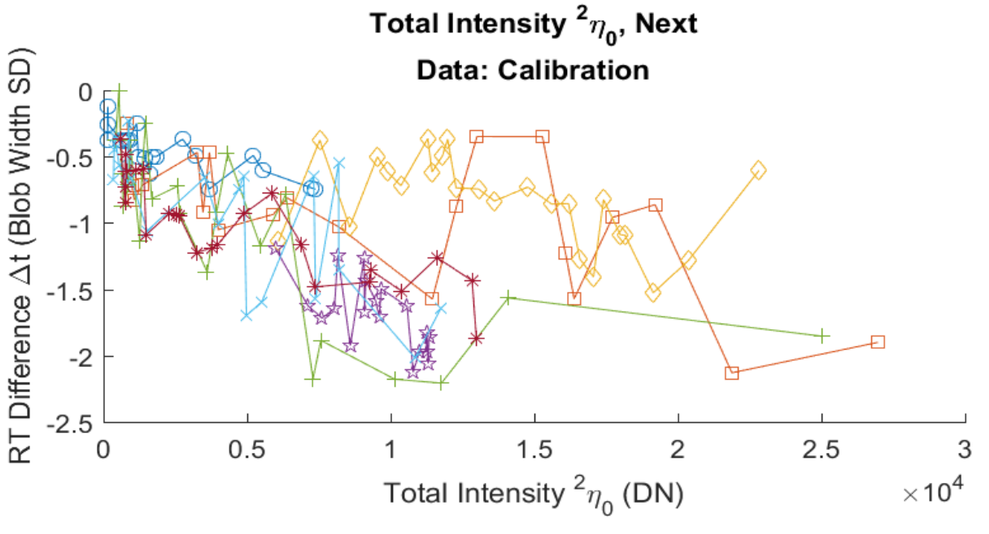
Increasing Pairs





Decreasing Pairs

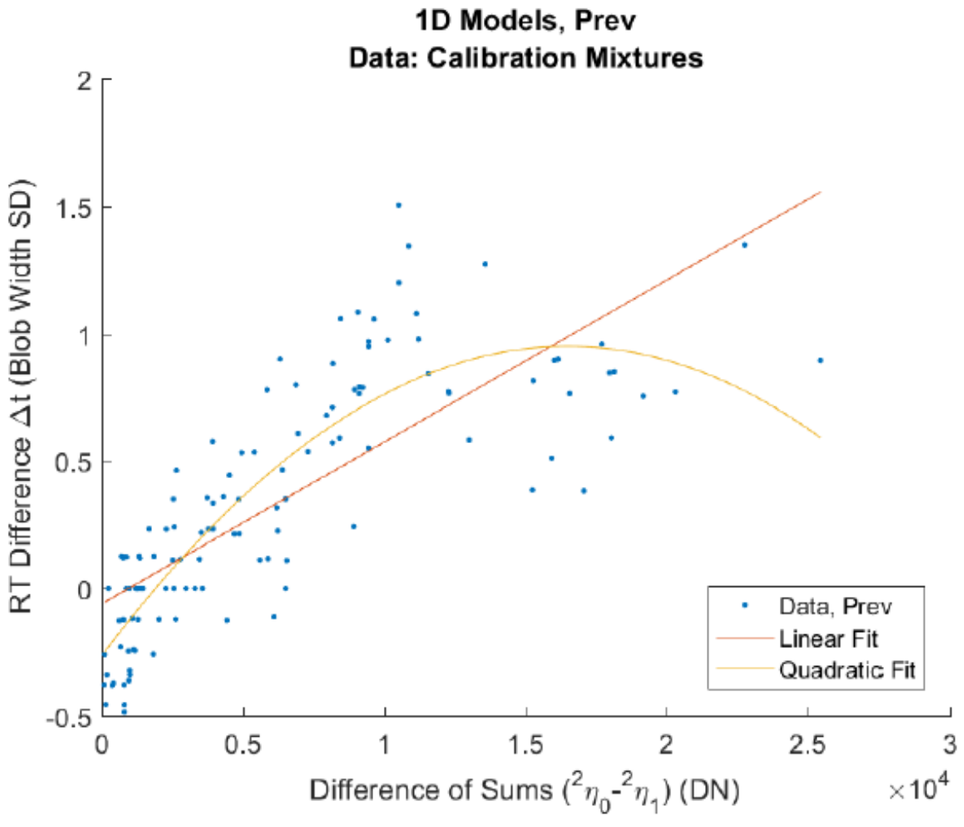




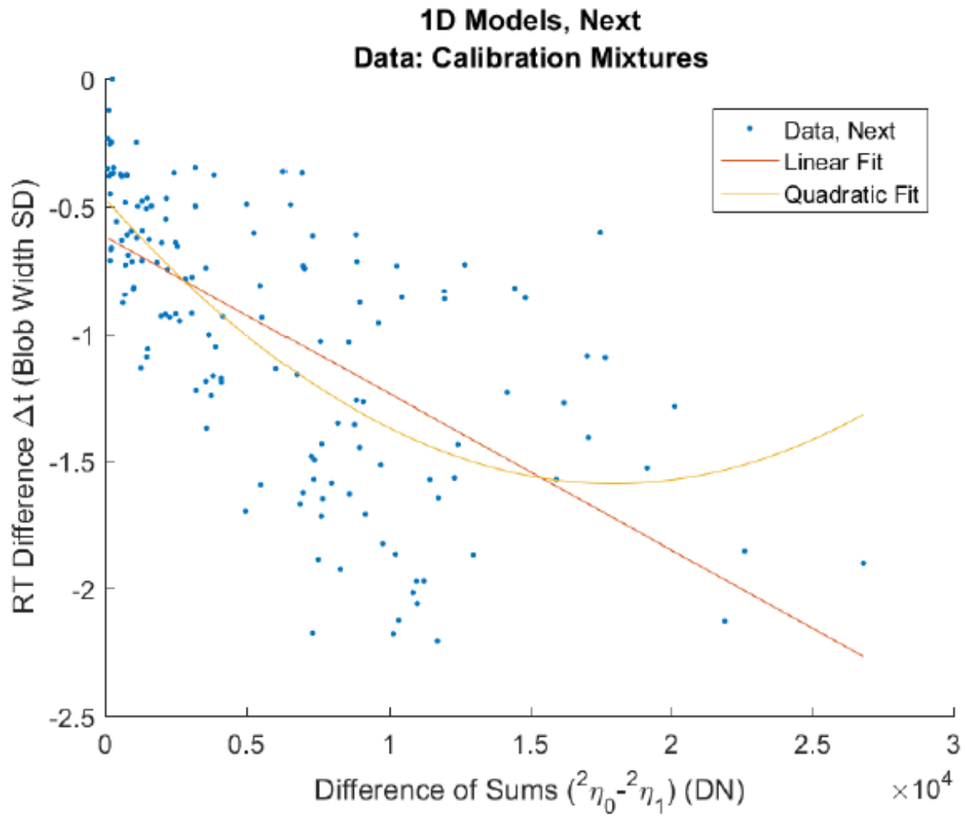
Secondary Independent Variables

- Investigated several secondary variables
- Difference of total intensities ($^2\eta_0 - {}^2\eta_1$) promising

Increasing Pairs



Decreasing Pairs



Linear, Univariate Models

- For linear, univariate models, computed RMS improvement
- Intensity difference performed well, with 60% decrease
  - $\Delta t_{est} = a \times ({}^{2}\eta_0 - {}^{2}\eta_1) + b$

	Calibration		Chem.Agts.		Temp.Ramp		Gas Press.		Average			%Impr.
Model	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**All**	**All**
None	0.57	1.12	0.38	0.51	0.44	0.57	0.84	1.17	0.56	0.84	0.70	
Mean	0.47	0.53	0.24	0.34	0.31	0.33	0.43	0.55	0.36	0.44	0.40	43%
$^{1}t_R$	0.46	0.52	0.22	0.32	0.31	0.30	0.42	0.55	0.35	0.42	0.39	45%
$^{2}t_R$	0.46	0.48	0.21	0.34	0.29	0.33	0.43	0.55	0.35	0.42	0.39	45%
$^{2}\eta_0$	0.35	0.46	0.21	0.34	0.29	0.30	0.42	0.52	0.32	0.40	0.36	48%
$^{2}\eta_1$	0.44	0.52	0.23	0.31	0.28	0.30	0.34	0.46	0.32	0.40	0.36	49%
$^{2}\sigma_0$	0.31	0.42	0.22	0.33	0.29	0.33	0.43	0.55	0.31	0.41	0.36	49%
$^{2}\sigma_1$	0.32	0.43	0.23	0.34	0.30	0.33	0.41	0.47	0.32	0.39	0.35	50%
$^{2}\gamma_0$	0.36	0.40	0.23	0.33	0.30	0.33	0.42	0.49	0.33	0.39	0.36	49%
$^{2}\gamma_1$	0.34	0.39	0.23	0.34	0.31	0.33	0.39	0.50	0.32	0.39	0.35	50%
$^{2}\eta_0$-$^{2}\eta_1$	0.29	0.41	0.18	0.29	0.23	0.21	0.28	0.35	0.25	0.31	0.28	60%
$^{2}\sigma_0$-$^{2}\sigma_1$	0.42	0.49	0.20	0.27	0.30	0.33	0.27	0.28	0.30	0.34	0.32	54%
$^{2}\gamma_0$-$^{2}\gamma_1$	0.30	0.31	0.24	0.32	0.30	0.32	0.37	0.43	0.30	0.34	0.32	54%

Linear, Bivariate Models

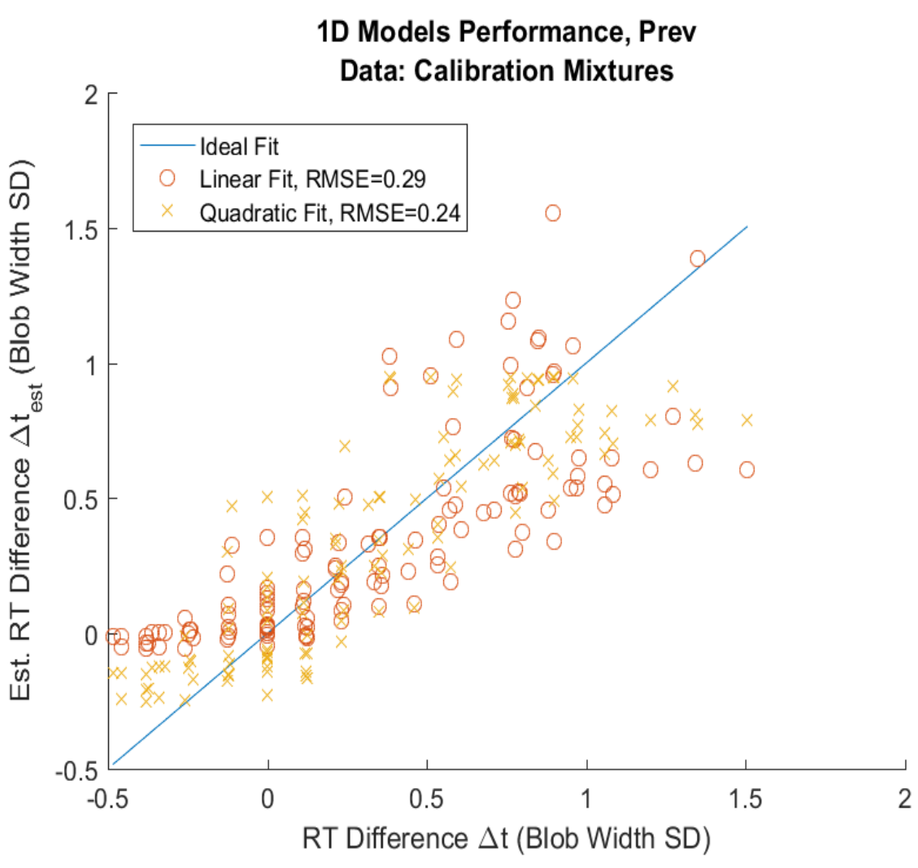
- For linear, bivariate models, computed RMS improvement
- Best bivariate models averaged 64-66% RMS reduction
  - Little better than 60% for best linear univariate model ($^2\eta_0 - {}^2\eta_1$)
- Additional complexity, more susceptible to over-fitting

	Calibration		Chem.Agts.		Temp.Ramp		Gas Press.		Average			%Impr.
Model	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**Prev**	**Next**	**All**	**All**
None	0.57	1.12	0.38	0.51	0.44	0.57	0.84	1.17	0.56	0.84	0.70	
$^2\eta_0, {}^2\eta_1$	0.27	0.39	0.17	0.25	0.20	0.19	0.19	0.30	0.21	0.28	0.24	65%
$^2t_R, {}^2\eta_0\text{-}{}^2\eta_1$	0.27	0.31	0.18	0.26	0.21	0.21	0.28	0.32	0.24	0.27	0.25	64%
$^2\sigma_1, {}^2\eta_0\text{-}{}^2\eta_1$	0.22	0.36	0.17	0.29	0.22	0.21	0.28	0.28	0.22	0.29	0.25	64%
$^2\gamma_0, {}^2\eta_0\text{-}{}^2\eta_1$	0.25	0.34	0.17	0.27	0.23	0.21	0.28	0.33	0.23	0.29	0.26	63%
$^2\gamma_1, {}^2\eta_0\text{-}{}^2\eta_1$	0.24	0.34	0.17	0.29	0.23	0.21	0.22	0.32	0.22	0.29	0.25	64%
$^2\eta_0\text{-}{}^2\eta_1, {}^2\sigma_0\text{-}{}^2\sigma_1$	0.23	0.36	0.18	0.26	0.23	0.21	0.22	0.24	0.22	0.27	0.24	66%
$^2\eta_0\text{-}{}^2\eta_1, {}^2\gamma_0\text{-}{}^2\gamma_1$	0.22	0.28	0.18	0.27	0.23	0.20	0.22	0.29	0.21	0.26	0.24	66%

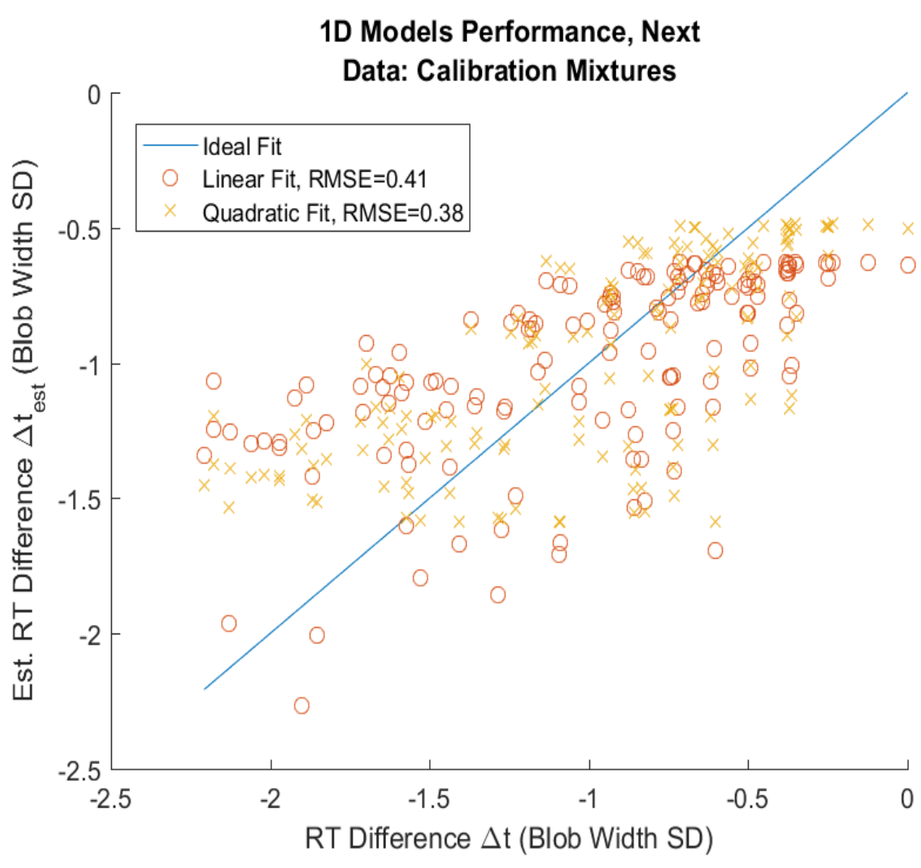
Higher-Order Models

- Investigated higher-order models (e.g., quadratic)
- Some, but not large improvements
- Same drawbacks as bivariate models: complexity & risk of overfitting

Increasing Pairs



Decreasing Pairs



Results Summary

- Linear, univariate model using intensity difference between peaks ($^2\eta_0 - {}^2\eta_1$) is good predictor
  - 60% improvement in RMS
- Bivariate and higher-order models provide only modest further improvement
  - May not be worth complexity & risk of over-fitting, depending on individual circumstances

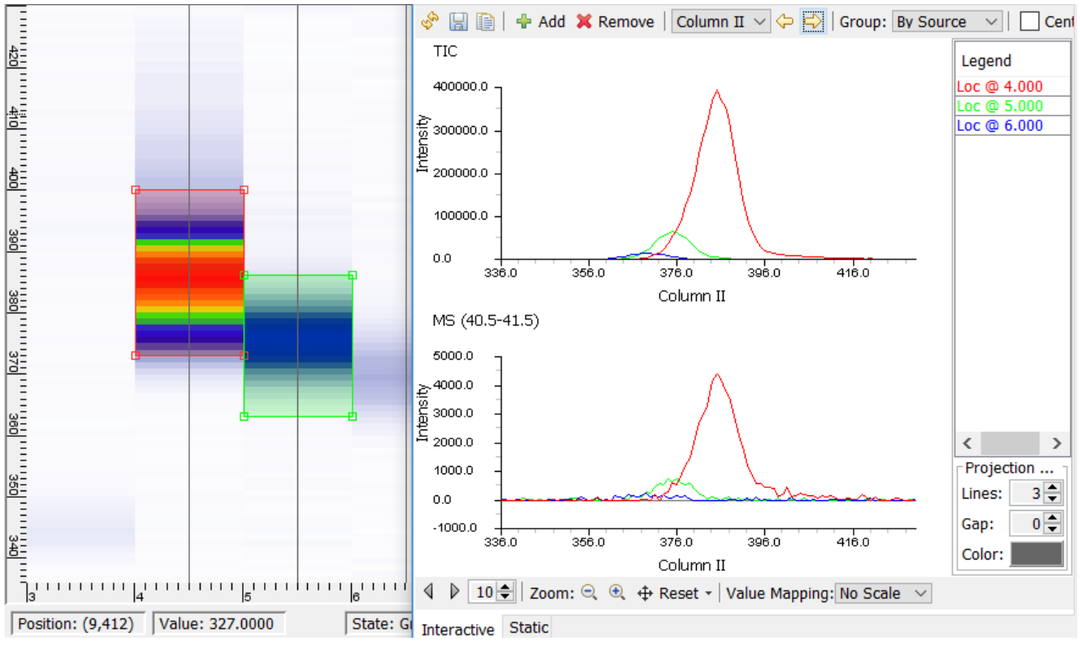
Dataset	Univariate Intensity Difference ($^2\eta_0 - {}^2\eta_1$)				Bivariate Intensities ($^2\eta_0, {}^2\eta_1$)			
	Increasing Pairs		Decreasing Pairs		Increasing Pairs		Decreasing Pairs	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
Calibration	0.29	0.24	0.41	0.38	0.27	0.21	0.39	0.33
Chem.Agts.	0.18	0.17	0.29	0.29	0.17	0.15	0.25	0.22
Temp.Ramp	0.23	0.22	0.21	0.20	0.20	0.18	0.19	0.16
Gas-Press.	0.28	0.27	0.35	0.34	0.19	0.17	0.30	0.27

Outline

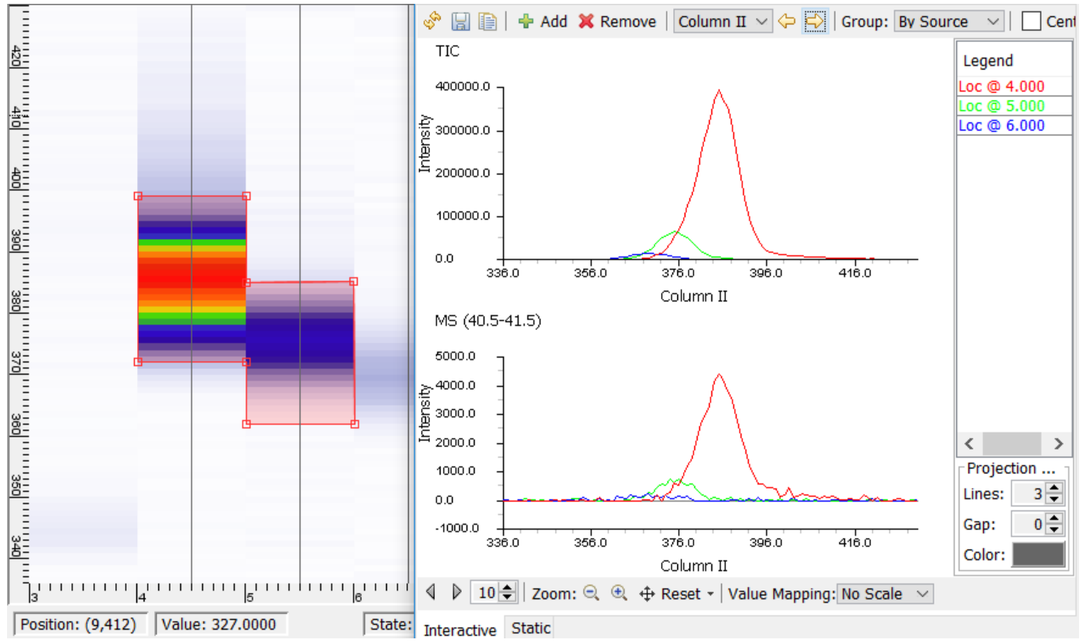
- Background
- Motivation
- Analysis
- Results
- Conclusion

Conclusion

- Analyzed RT shift in blobs for multiple datasets, and it's relationship to peak/blob characteristics
- Univariate, linear model using intensity difference ($^2\eta_0 - {}^2\eta_1$) is simple, but effective predictor
- Promising for applications such as ion blob detection
  - Work in progress, testing & refining with more complex data



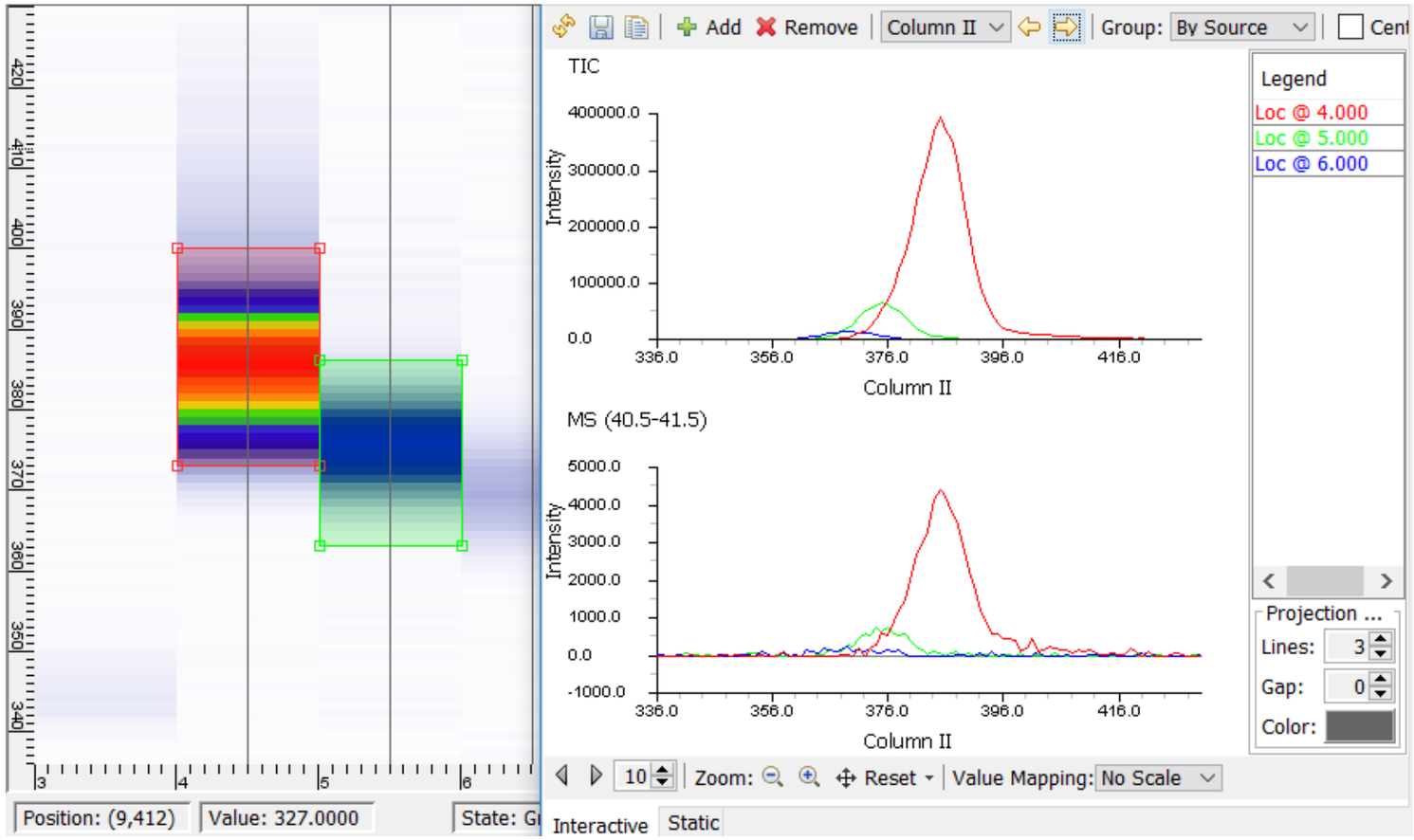
a) No prediction model



b) Using prediction model

Application: Ion Blob Detection

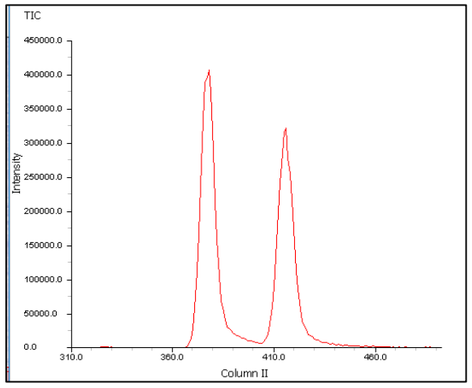
- **Goal**: Detect 2D blobs for individual ions in GCxGC-MS
- Use linear (${}^2\eta_0 - {}^2\eta_1$) model to predict RT shift during ion blob detection
- Avoid detecting multiple blobs when large shift in RT
- Example: should be single compound blob



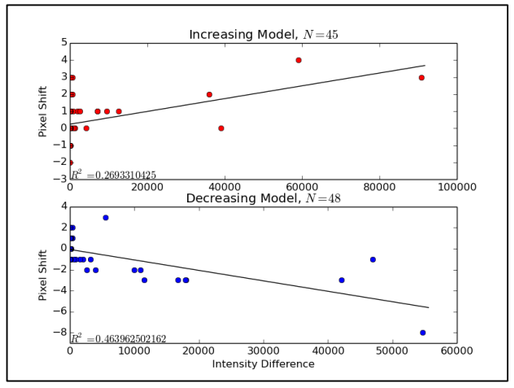
Application: Ion Blob Detection (2)

- Implemented as part of modified watershed algorithm:
  - Perform ^{2}D peak detection
  - Fit increasing/decreasing linear models based on selected ^{2}D peak-pairs
  - Detect 2D blobs based on model

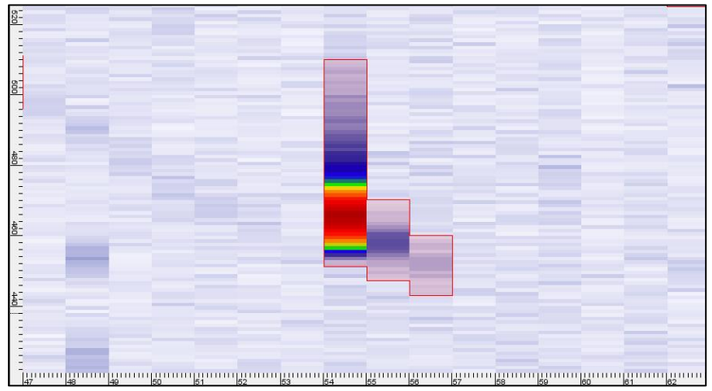
1) ^{2}D peak detection



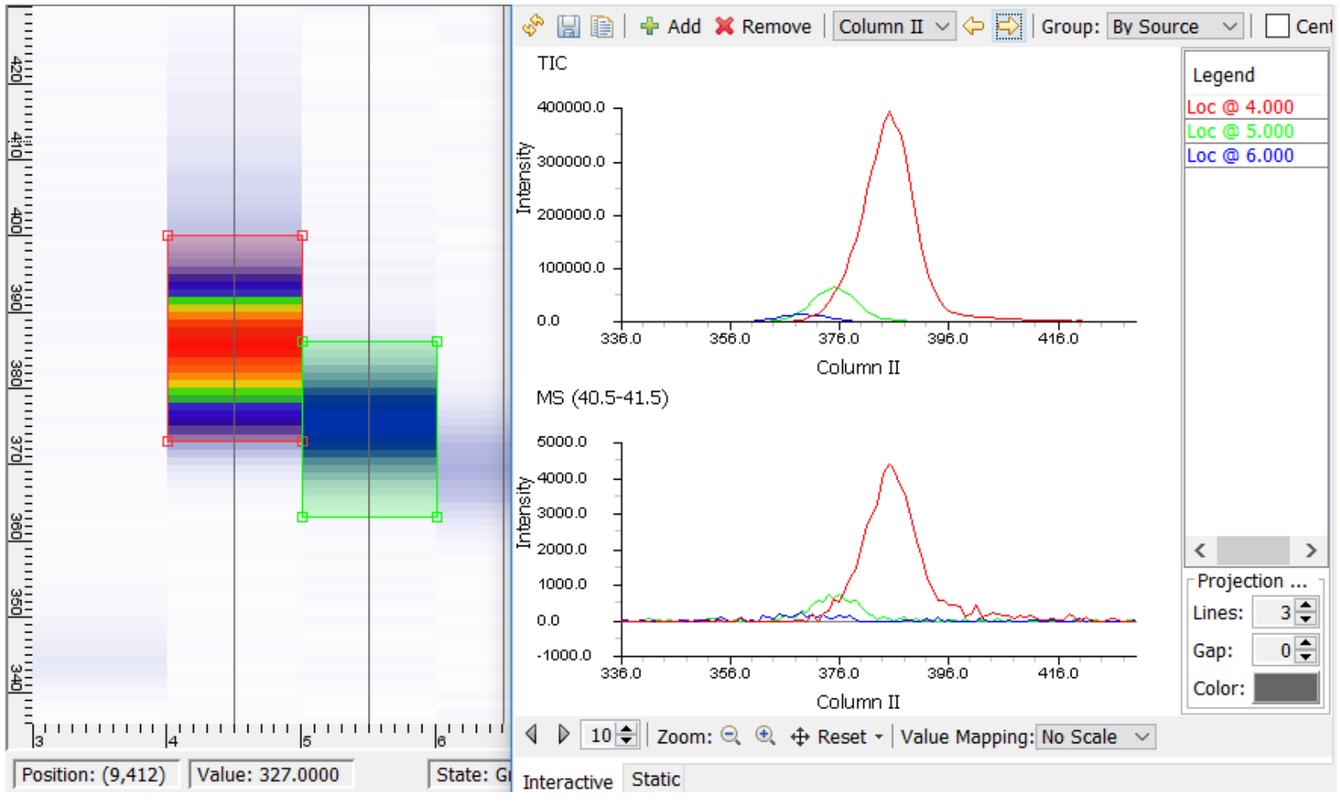
2) Fitting models



3) 2D blob detection



a) No prediction model



b) Using prediction model