Visualization of Two-way, Three-way and Higher Order Data Sets

Barry M. Wise

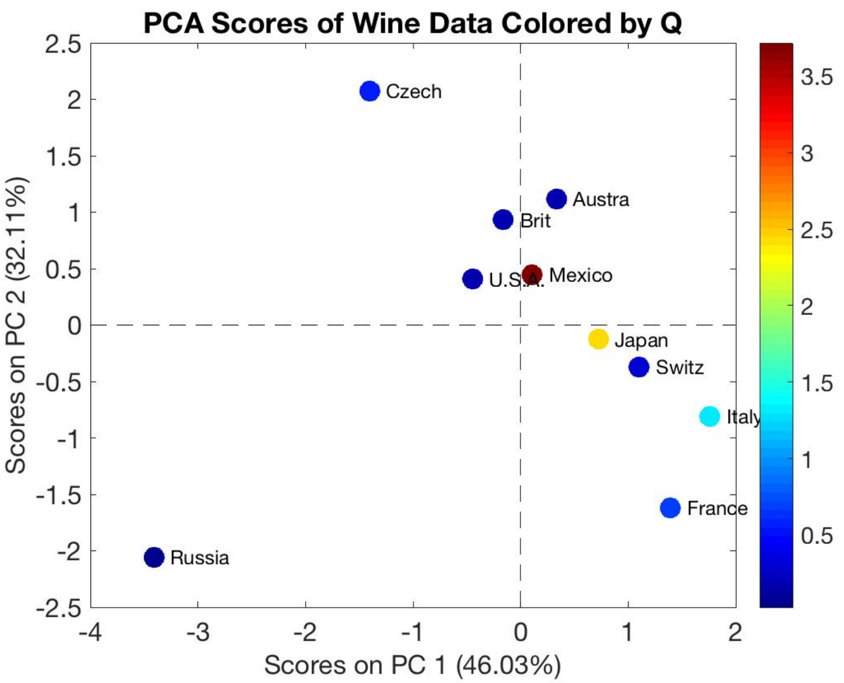
EIGENVECTOR RESEARCH, INC.

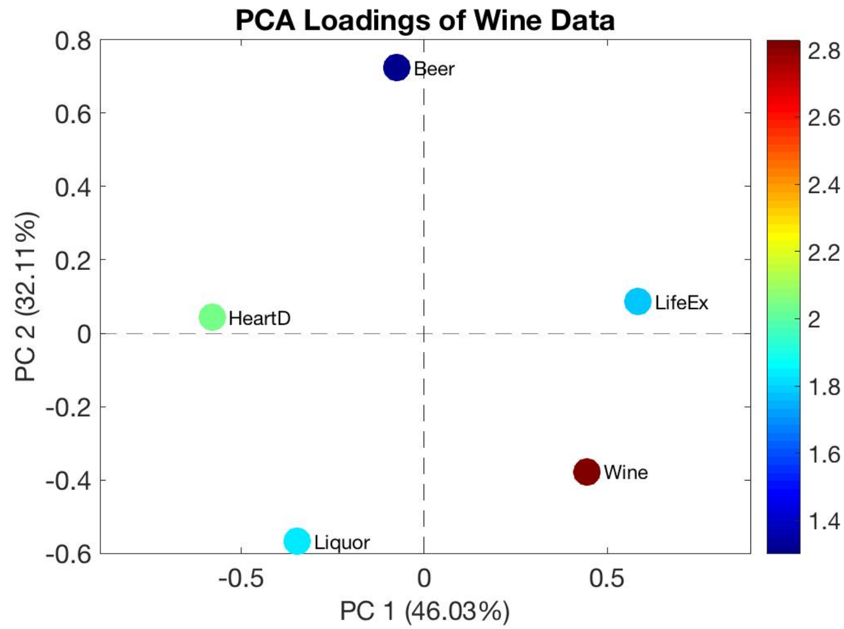
Data Order

- Data Order defined by the number of dimensions in which it is logically arranged
- Two-way: conventional data tables
  - *Samples* by *variables*
- Three-way: data cubes
  - EEM: *Samples* by *excitation* by *emission*
  - GC-MS: *Samples* by *retention* by *mass number*
  - Hyperspectral: *x* by *y* by *spectra*
- Four-way: series of data cubes
  - SIMS depth profiling: *x* by *y* by *z* by *spectra*
- Five-way: array of data cubes
  - EEM images: *sample* by *x* by *y* by *excitation* by *emission*
- Etc.

EIGENVECTOR RESEARCH INCORPORATED

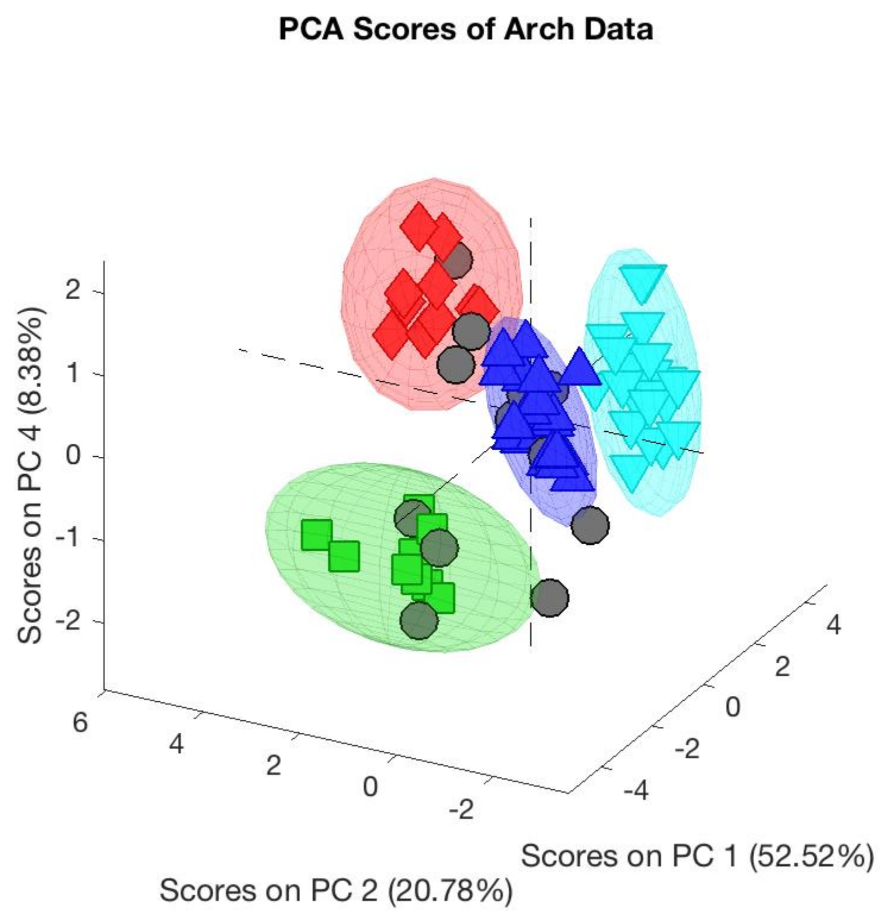
Visualization Progress





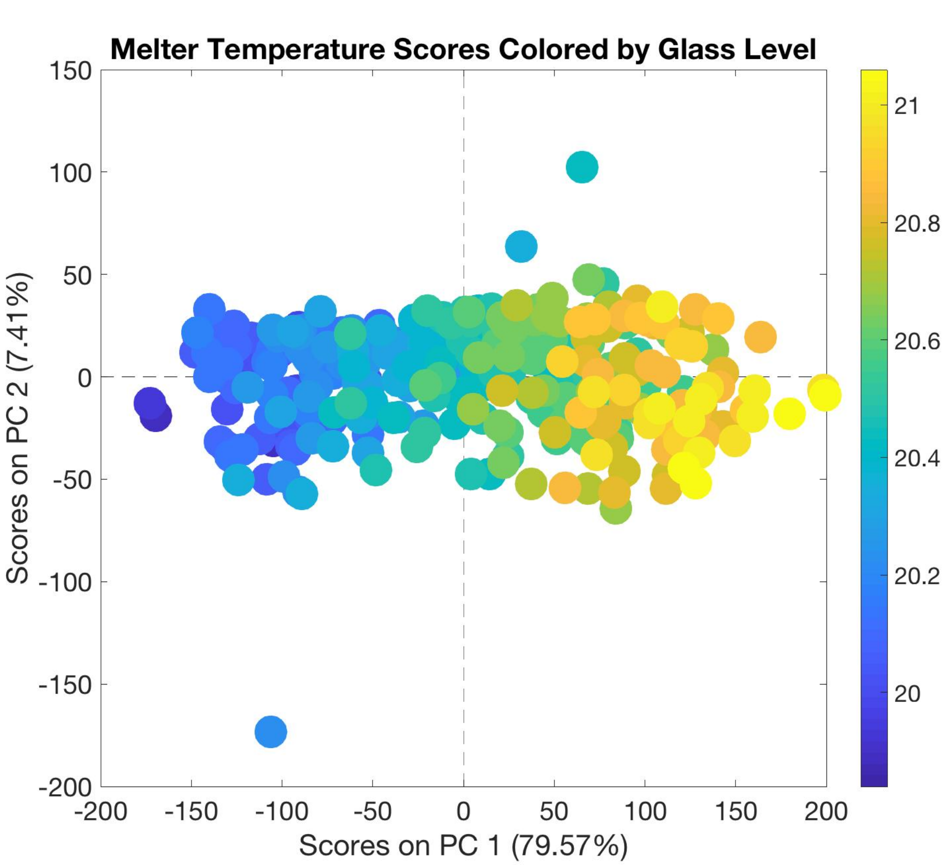
EIGENVECTOR RESEARCH INCORPORATED

Continued Refinement

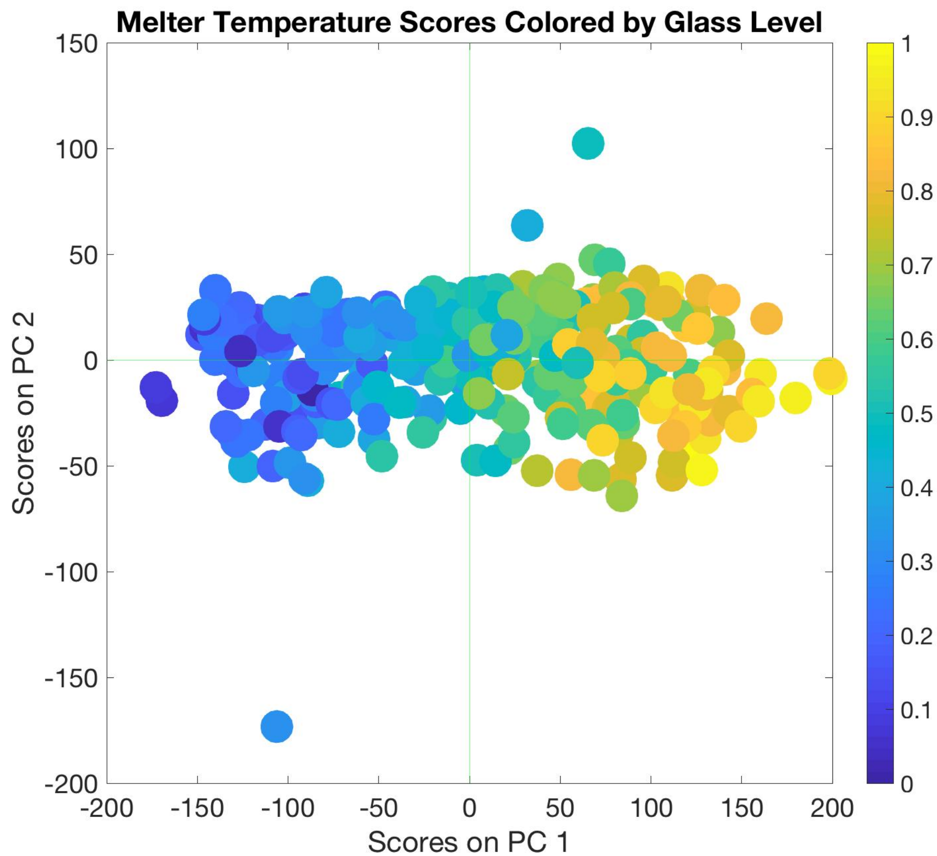


EIGENVECTOR RESEARCH INCORPORATED

Issues With Larger Data Sets



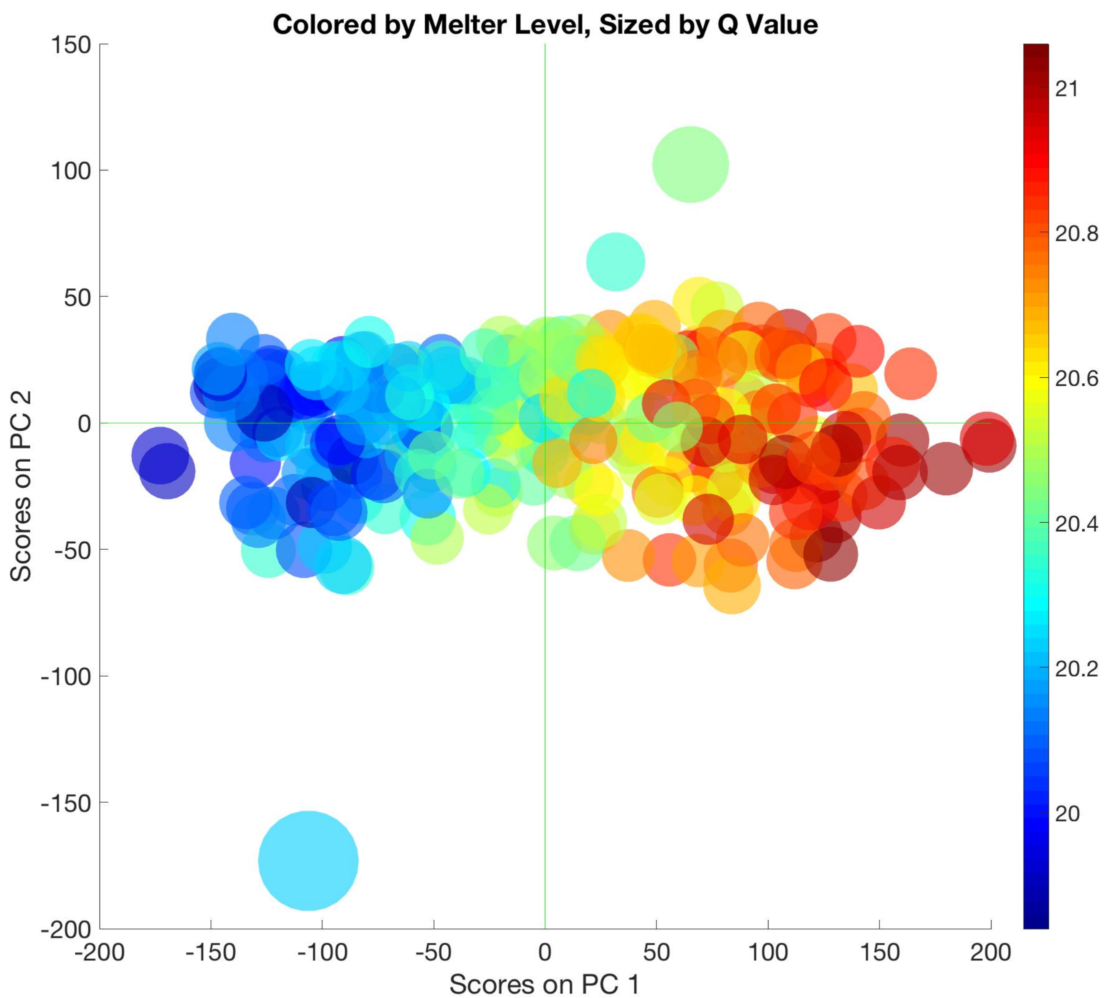
Colorby Layered



Colorby Unsorted

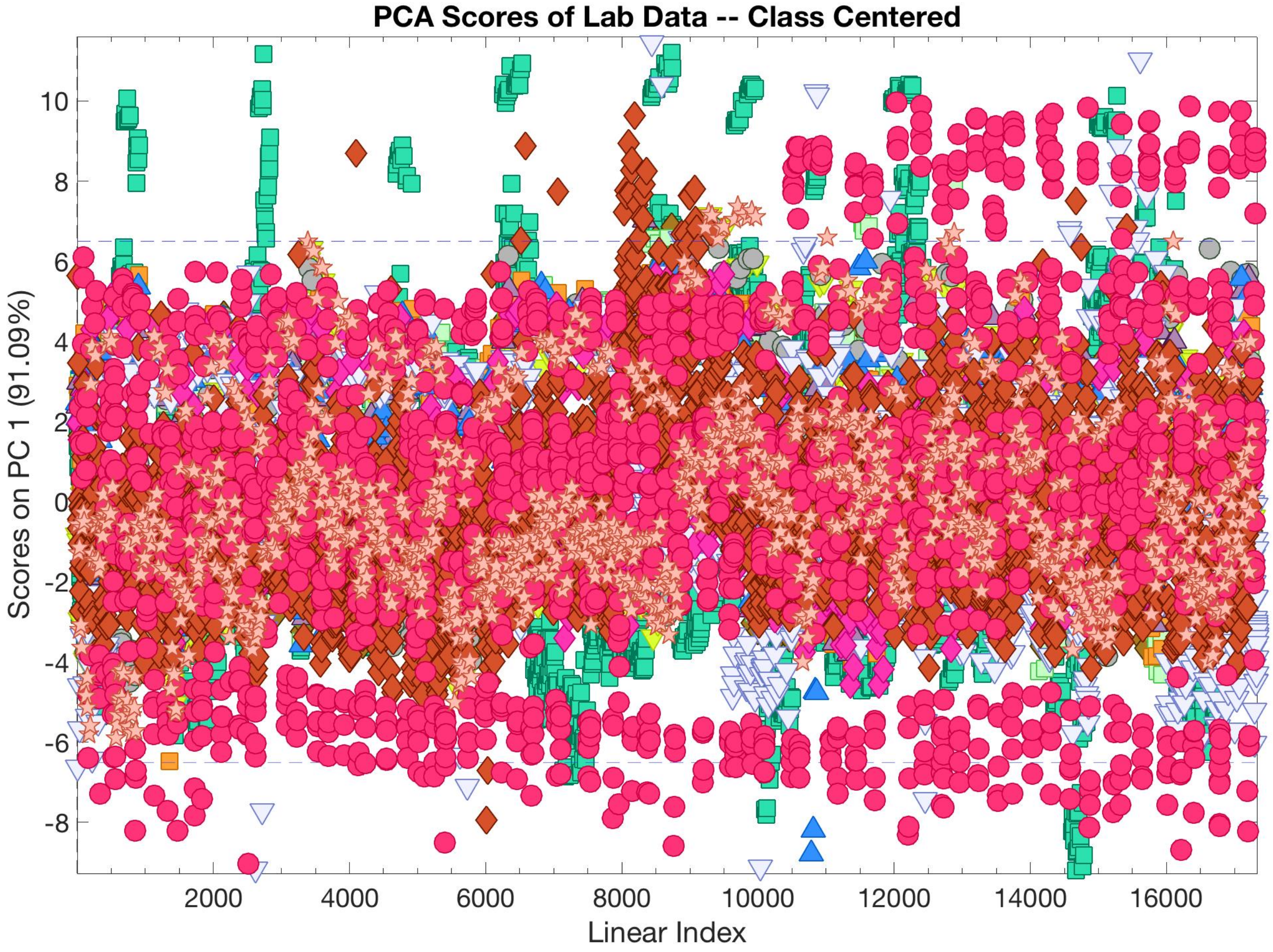
EIGENVECTOR RESEARCH INCORPORATED

Translucent Data Points? Size?



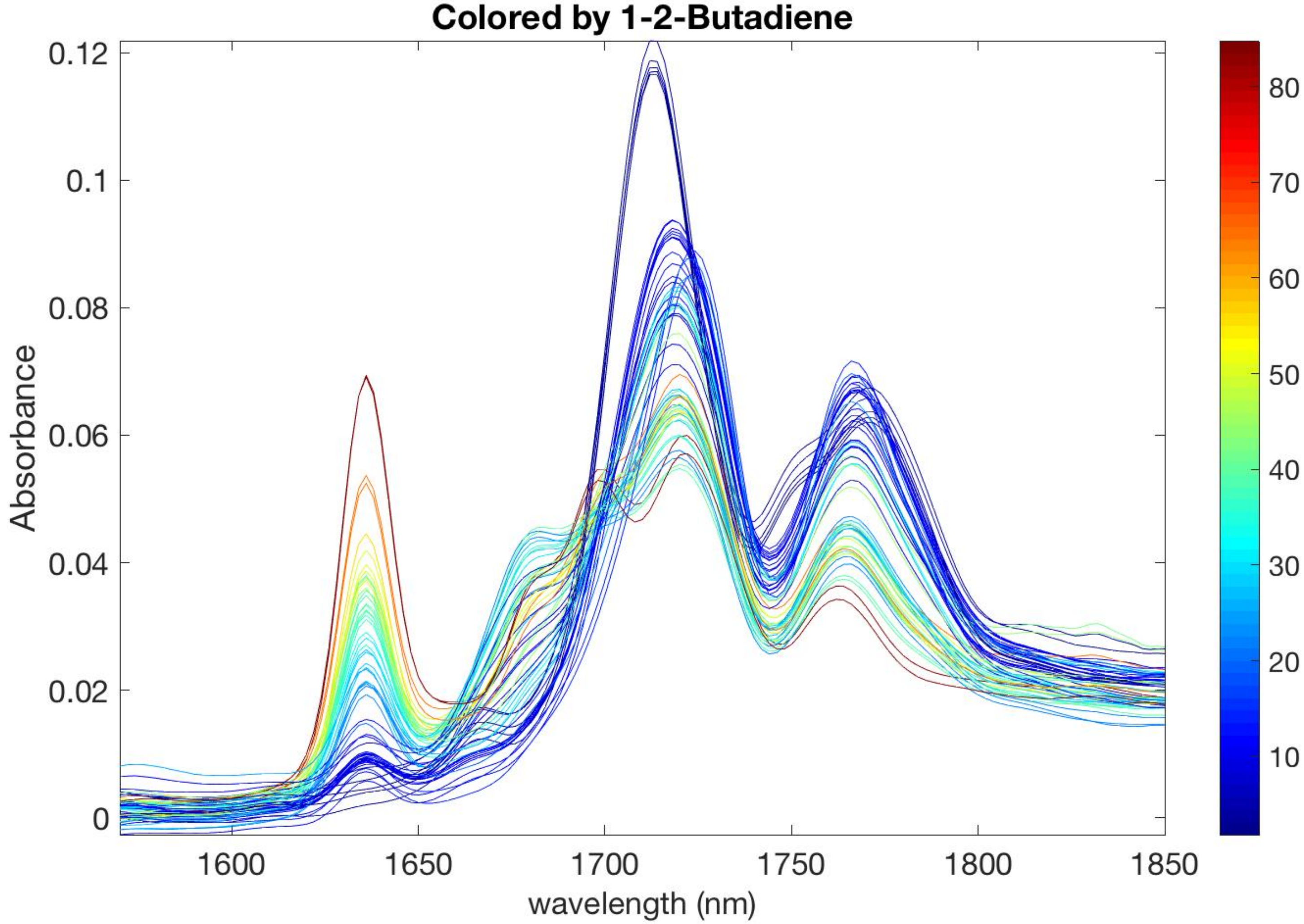
EIGENVECTOR RESEARCH INCORPORATED

And There Are Limits



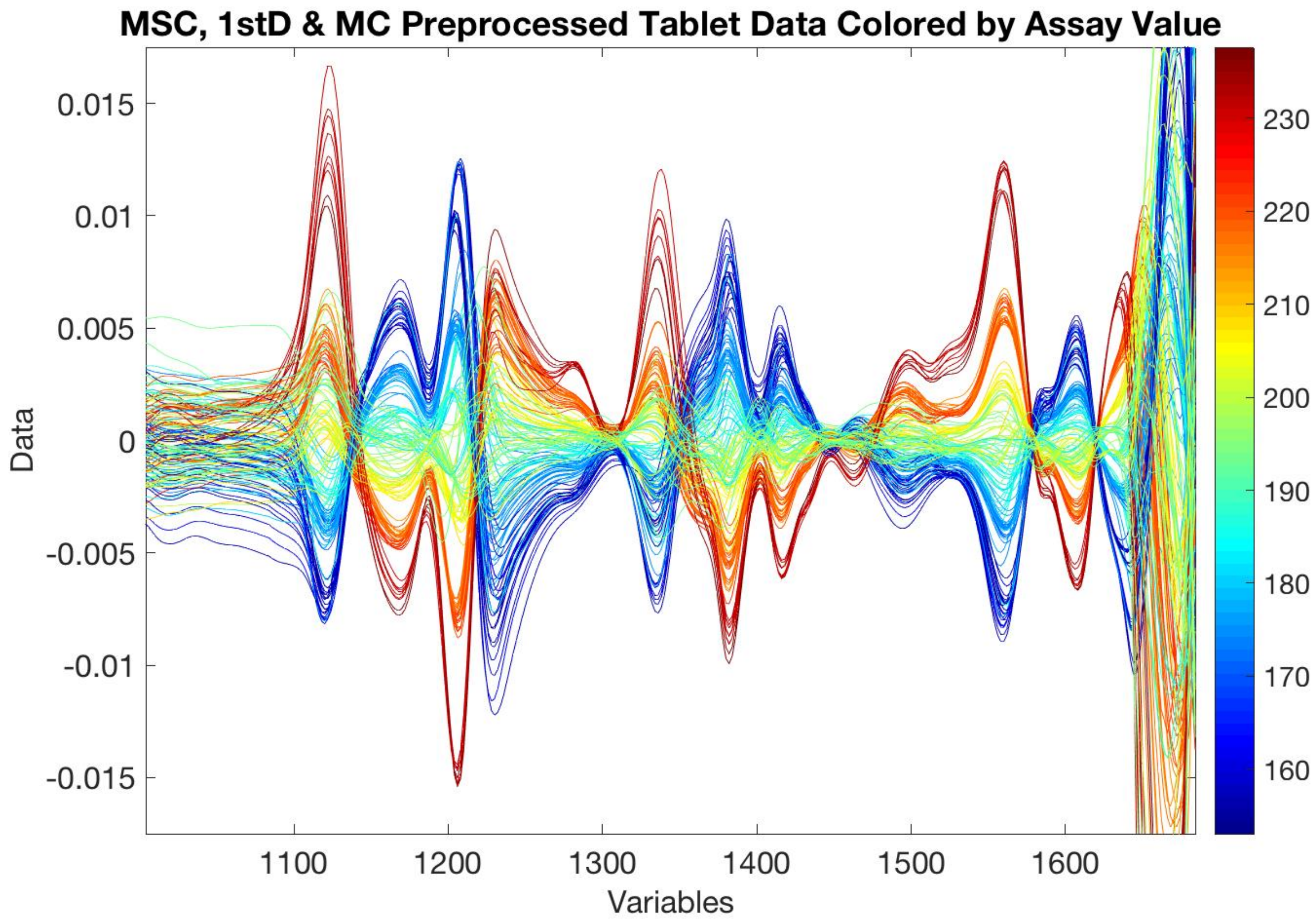
EIGENVECTOR RESEARCH INCORPORATED

Visualizing Raw Data

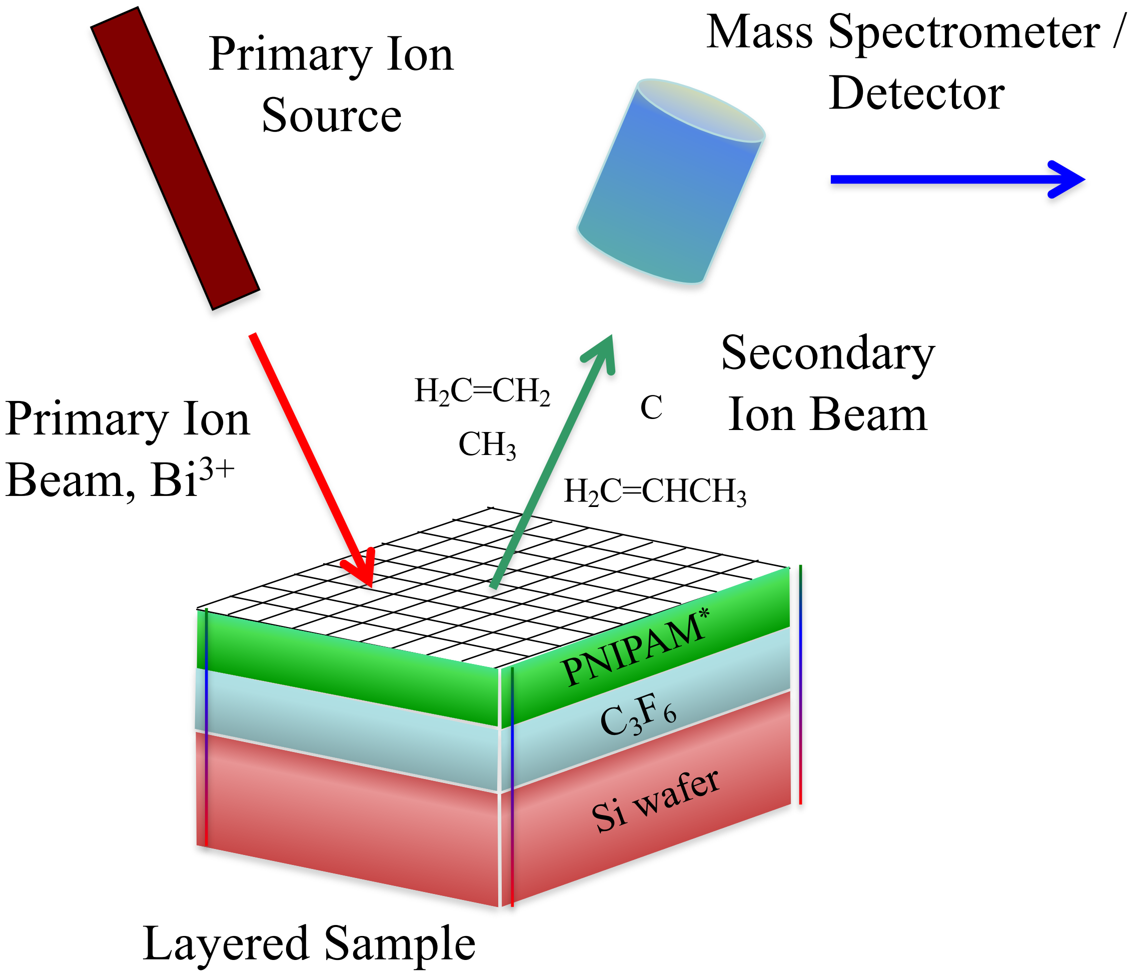


EIGENVECTOR RESEARCH INCORPORATED

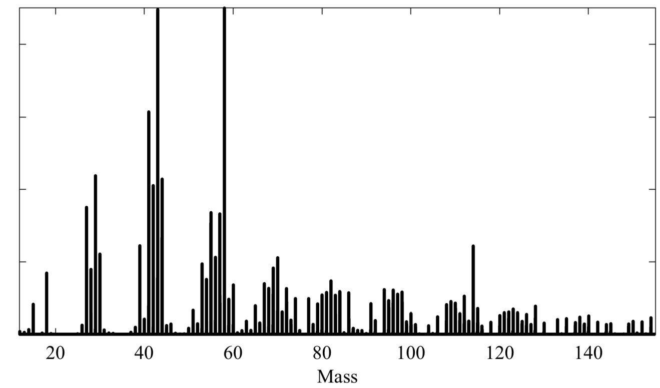
Visualizing Preprocessing



Secondary Ion Mass Spectrometry (SIMS)



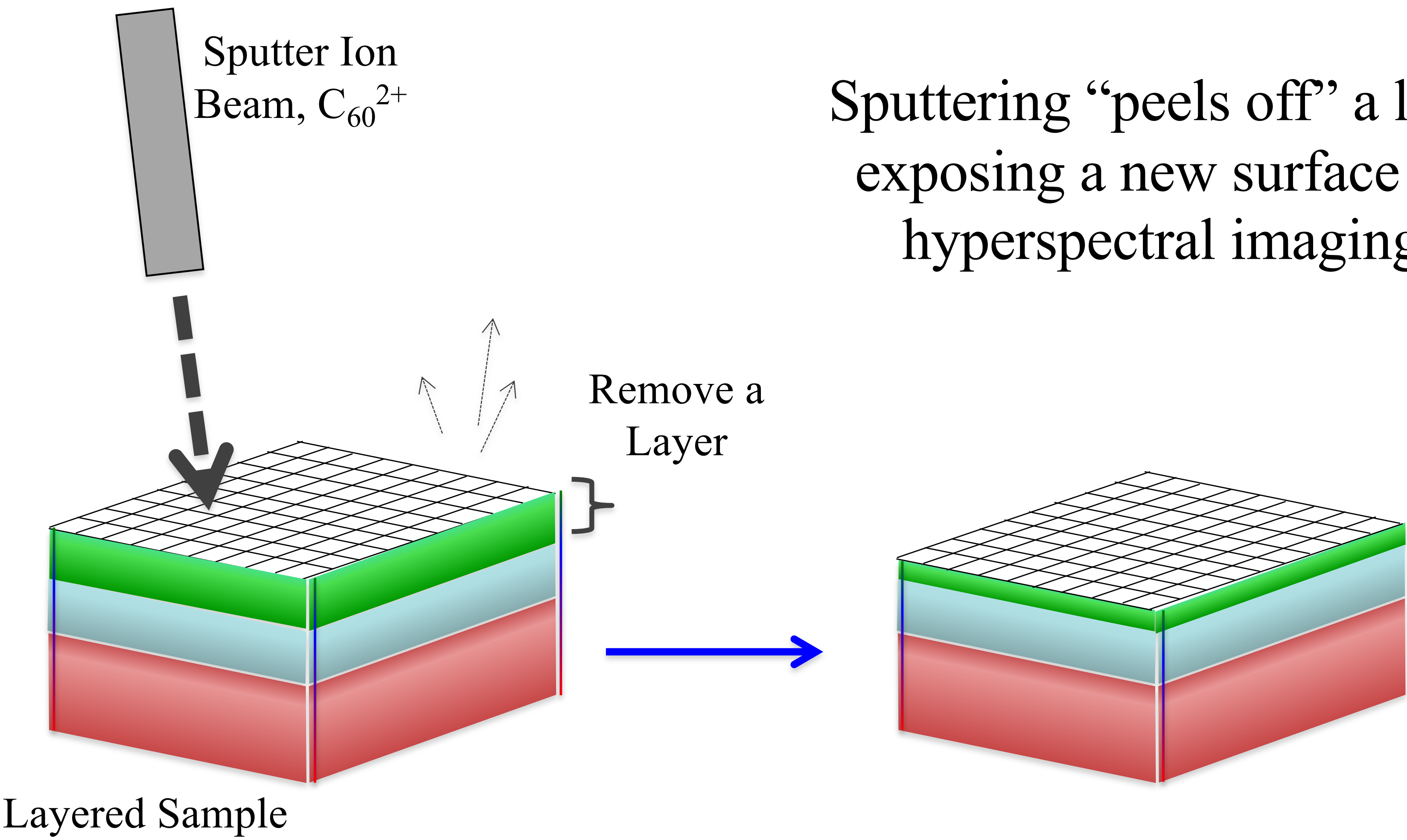
*Poly(N-iopropylacrylamide)



The primary ion source is moved over the sample surface in a pseudo-random pattern to obtain a hyperspectral image.

EIGENVECTOR RESEARCH INCORPORATED

Sputtering → Depth Profiling



Sputtering "peels off" a layer exposing a new surface for hyperspectral imaging.

EIGENVECTOR RESEARCH INCORPORATED

ToF-SIMS Imaging & Depth Profiling

- The primary beam is scanned over the surface
  - mass channel spectrum (time of flight) at each pixel
  - 256x256 hyperspectral image of the surface
- Depth profiling is achieved by sputtering
  - multiple hyperspectral images at 50 different depths (~200 nm depth profile)
- 256x256x406+x50 reduced to 85x85x300x50
  - use mean of 3x3 windows, remove edge pixels, remove highest mass channels

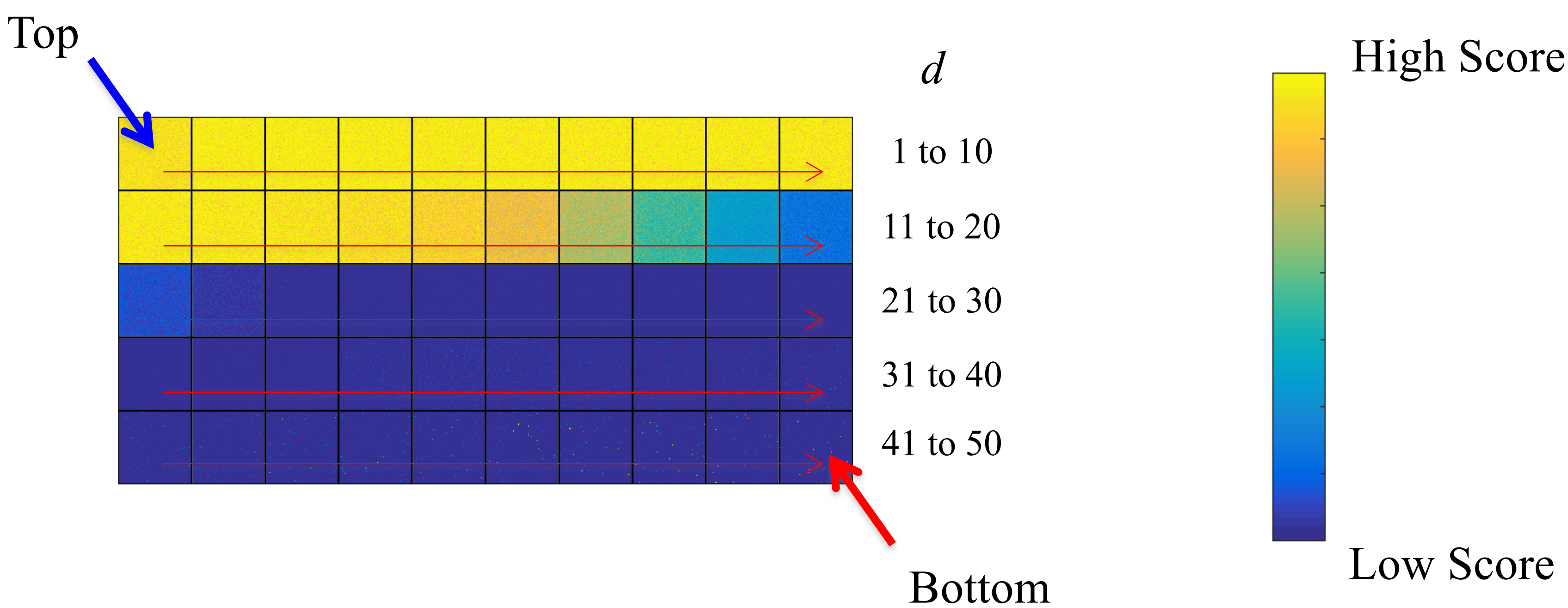
+406 peaks were selected and integrated across all mass spectr

EIGENVECTOR RESEARCH INCORPORATED

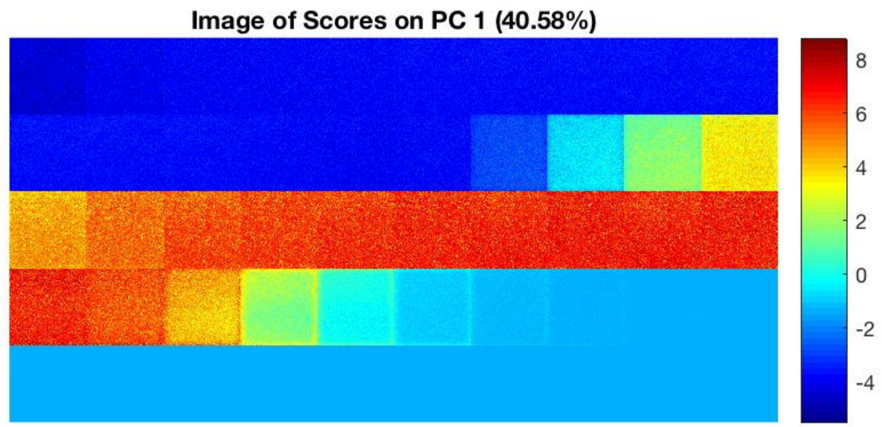
Data Analysis

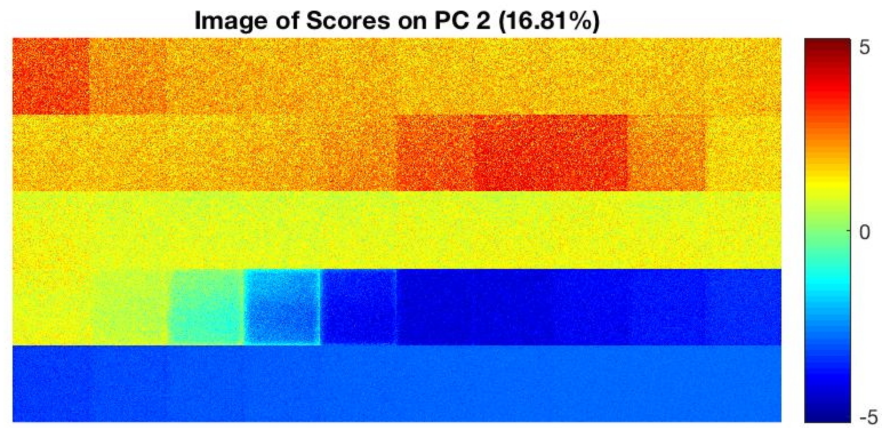
- Principal Components Analysis (PCA)
  - orthogonal scores / loadings
- Multivariate Curve Resolution (MCR)
  - attempt to obtain pure component, non-negative scores and factors that are more physically interpretable
- Use Poisson scaling
  - allows lower signal (higher mass channels) to influence the model
- Parallel Factor Analysis (PARAFAC)

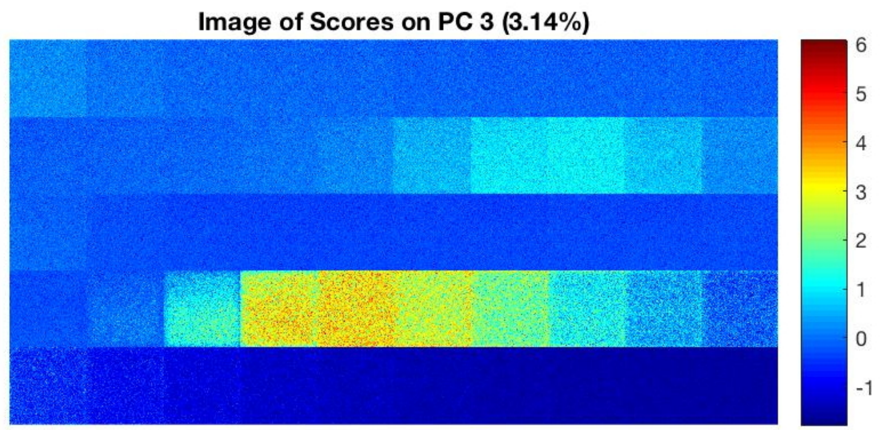
Example Image

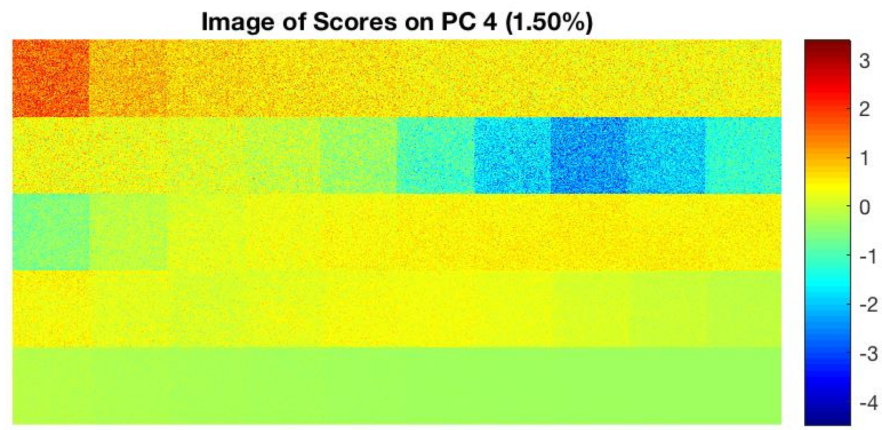


PCA Results Tiled Scores

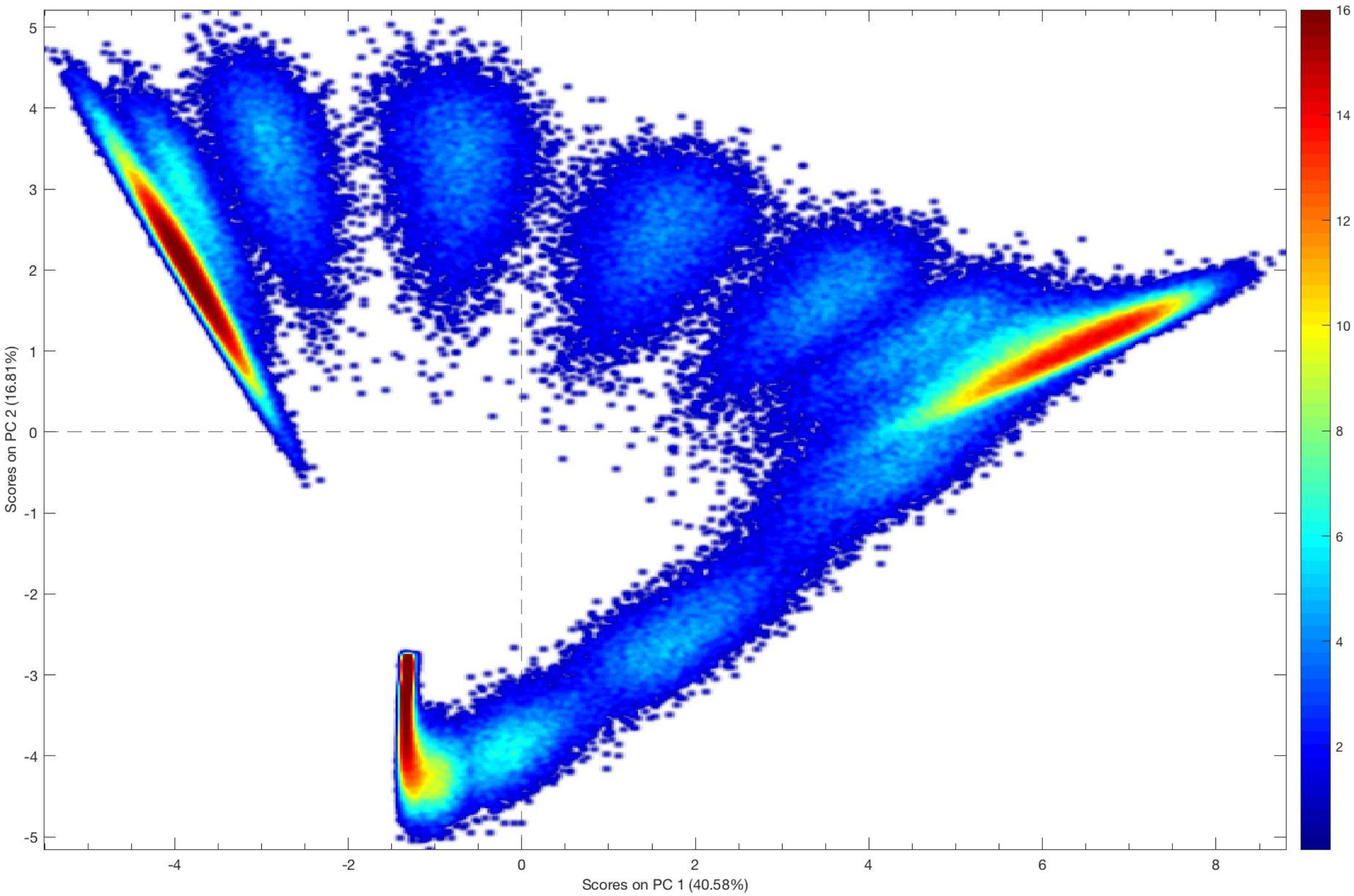




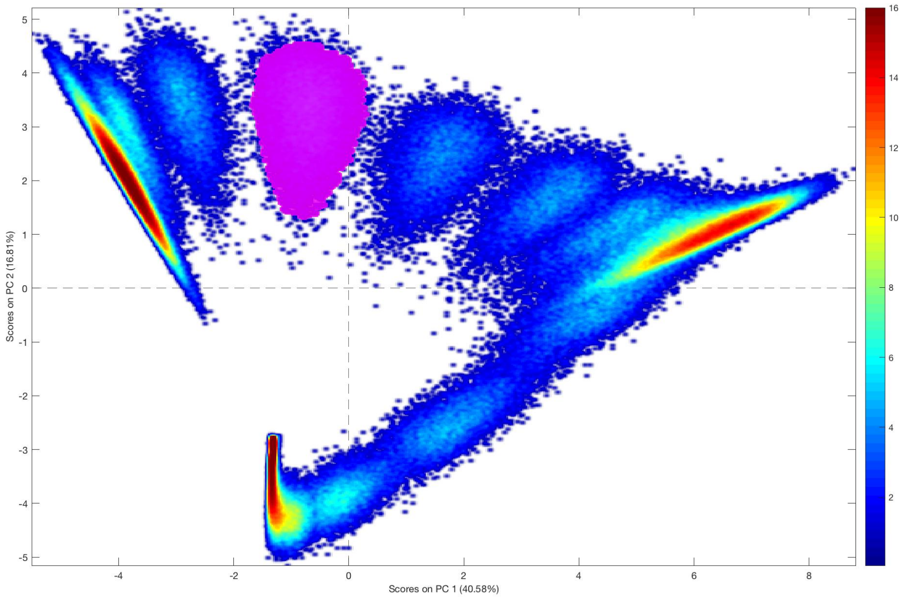


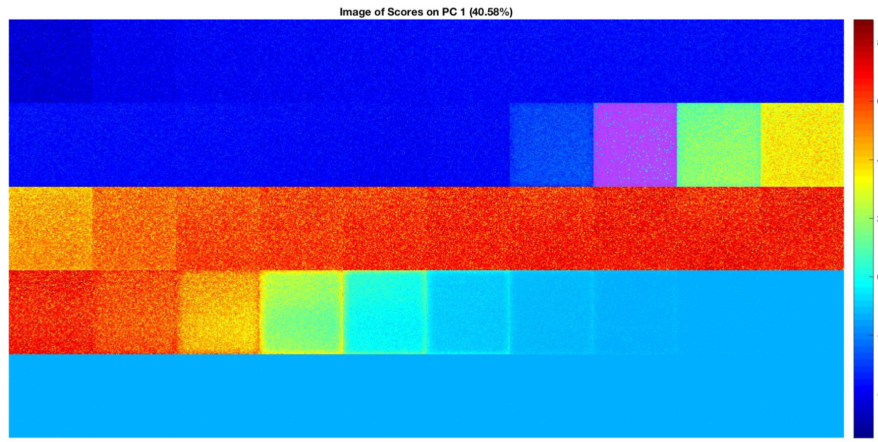


Score-Score Plots



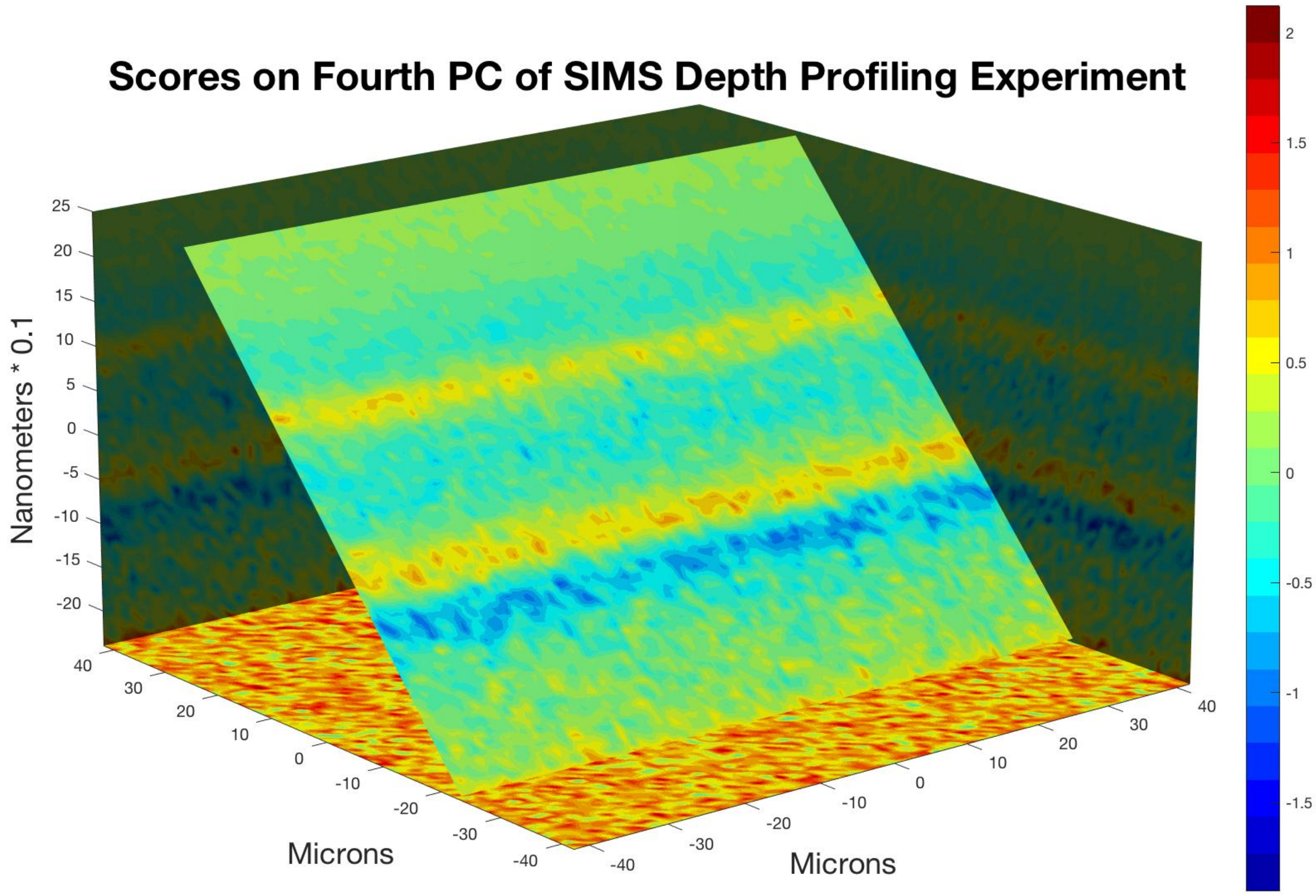
Linking



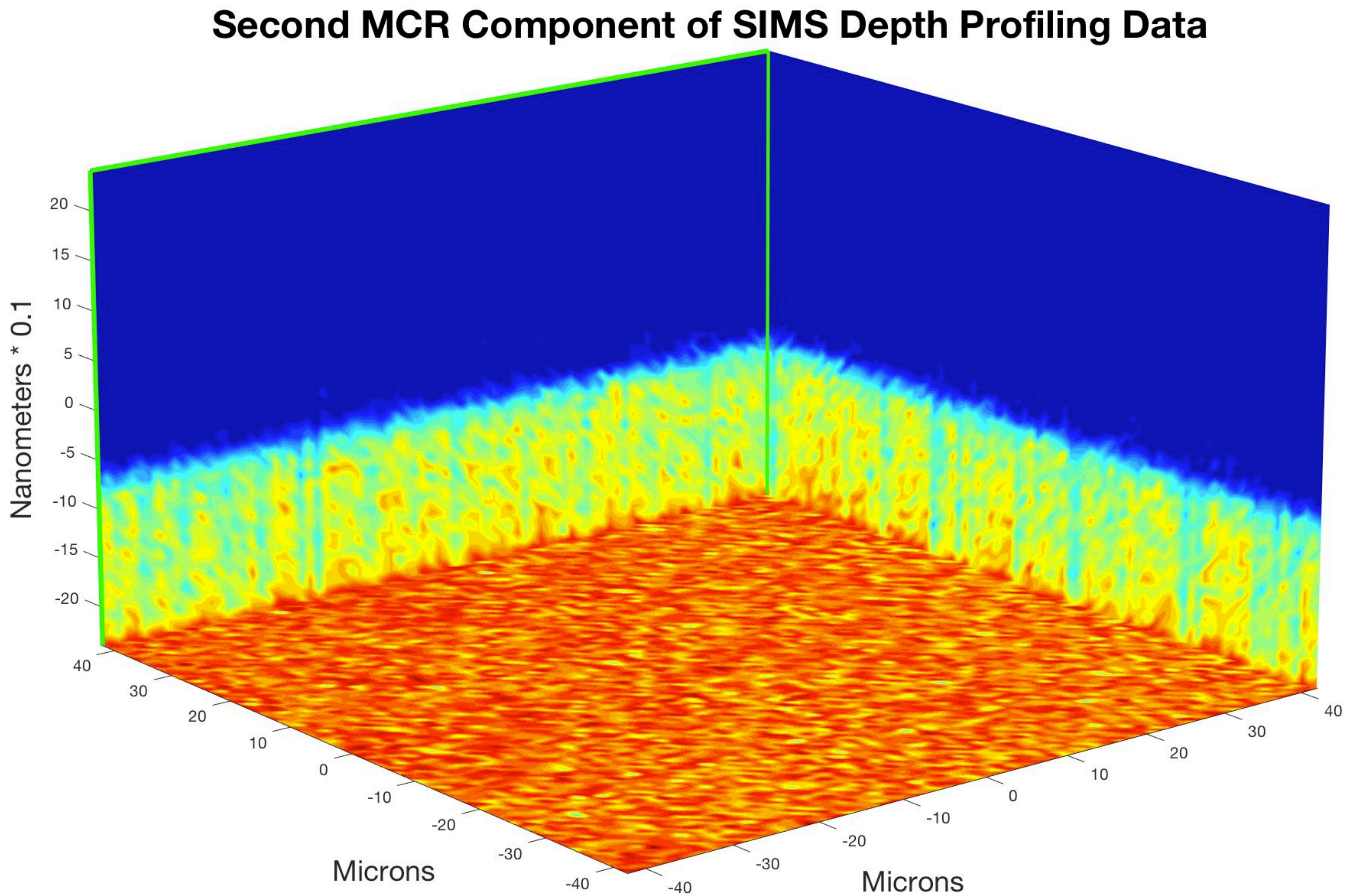


EIGENVECTOR RESEARCH INCORPORATED

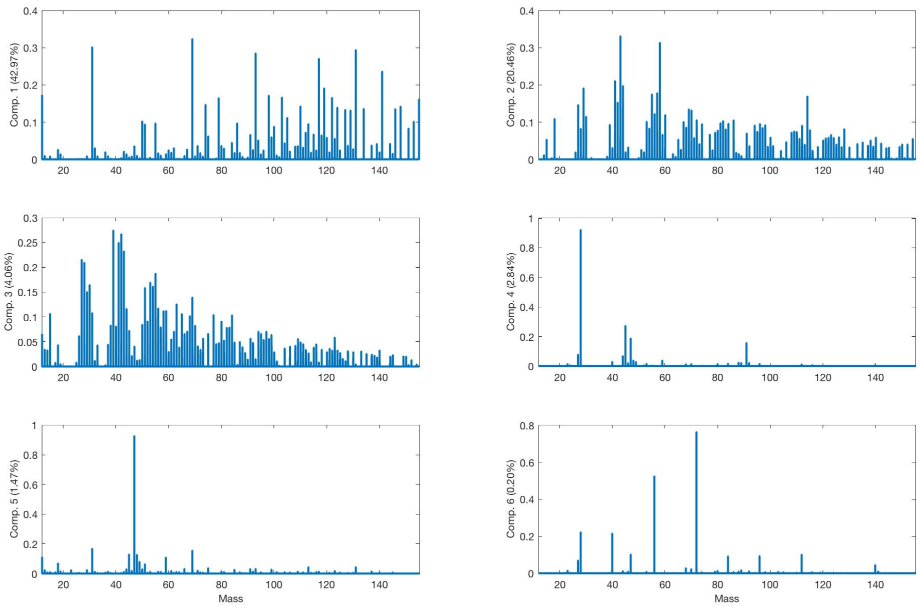
Visualizing 3-D Volume

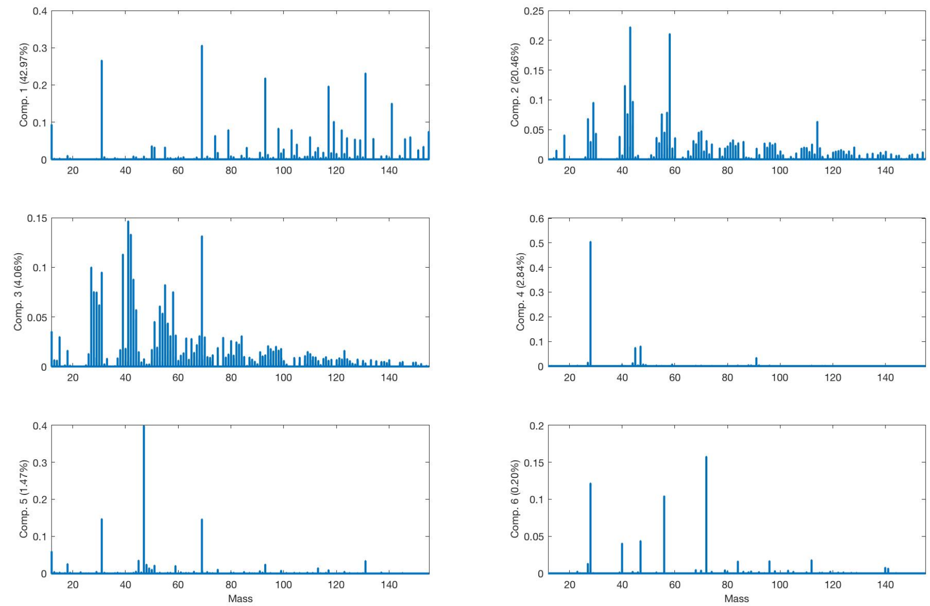


EIGENVECTOR RESEARCH INCORPORATED


Second MCR Component of SIMS Depth Profiling Data
Nanometers * 0.1
20
15
10
5
0
-5
-10
-15
-20
Microns
40
30
20
10
0
-10
-20
-30
-40
Microns
-40
-30
-20
-10
0
10
20
30
40

Undo Preprocessing





As recovered

Preprocessing undone

EIGENVECTOR RESEARCH INCORPORATED

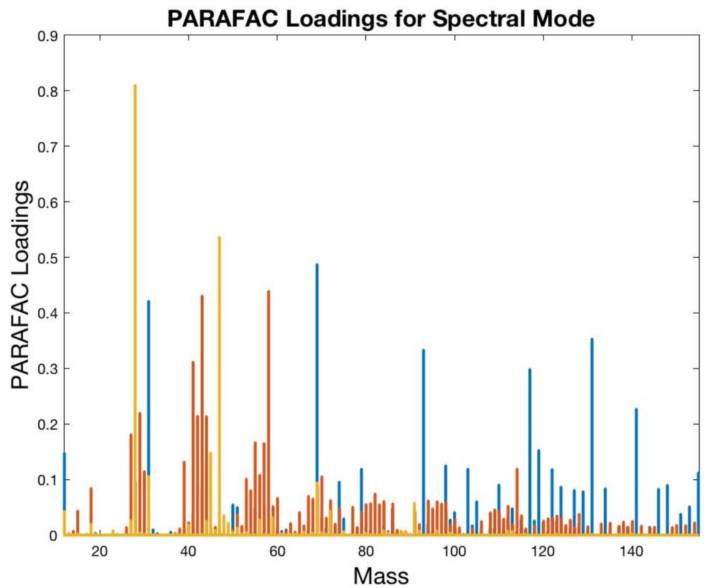
Another take: PARAFAC

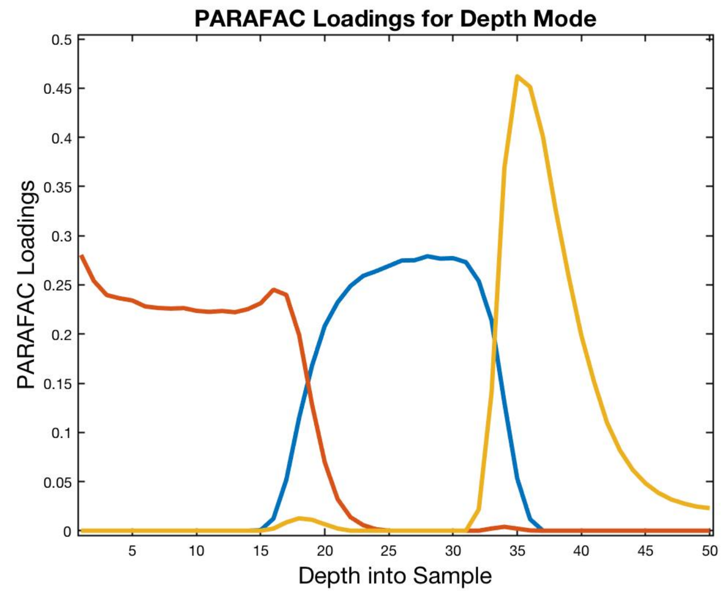
- Unfold 4-way data array in pixel (x-y) modes
  - 7225 pixels by 300 mass channels by 50 layers
- Use Parallel Factor Analysis to model data as outer product of the three modes
- Fold pixel mode back up to make images
- Used non-negativity constraint

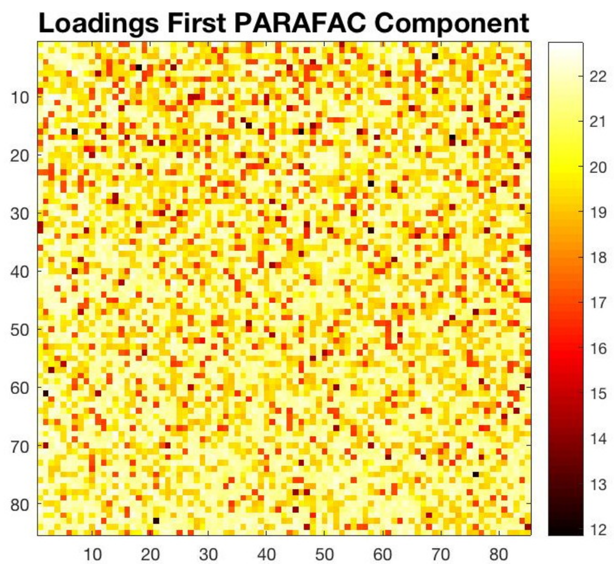
EIGENVECTOR RESEARCH INCORPORATED

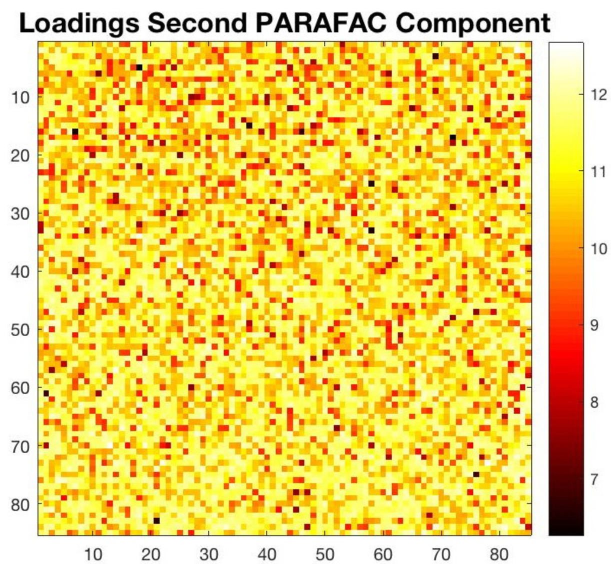
PARAFAC Model

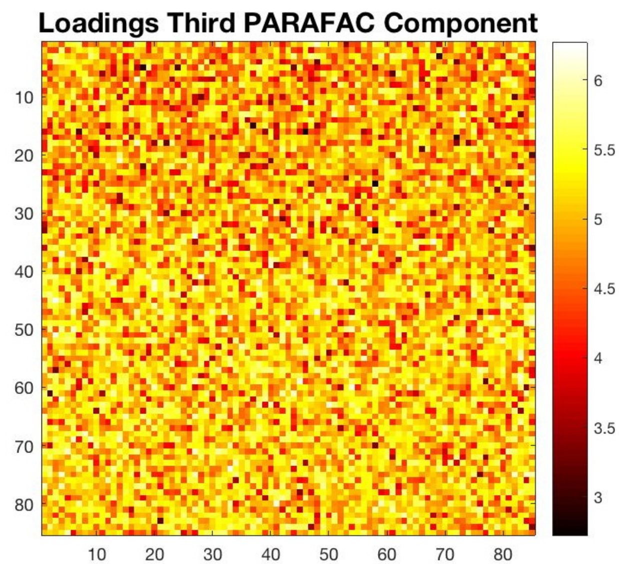
88.1% Variance Captured!



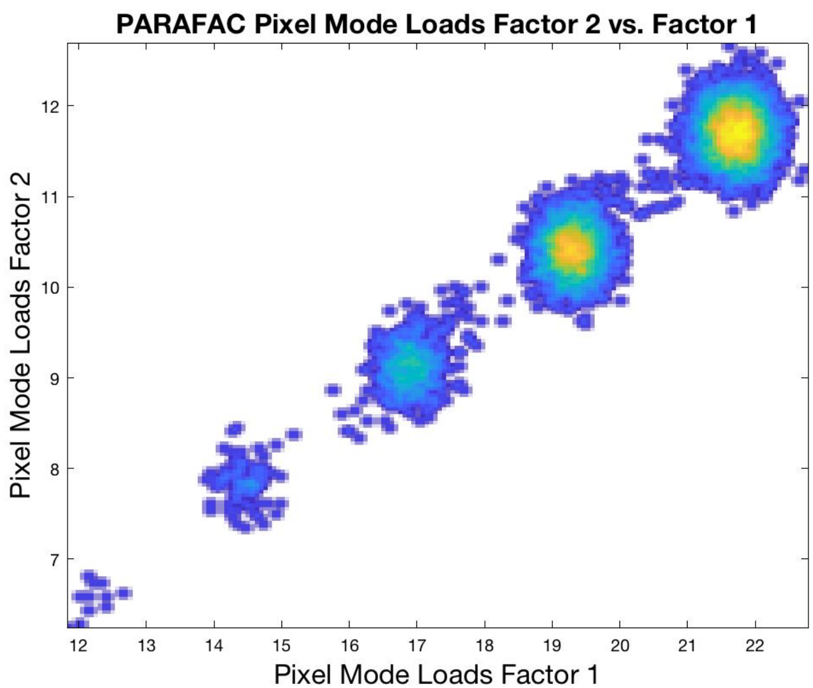


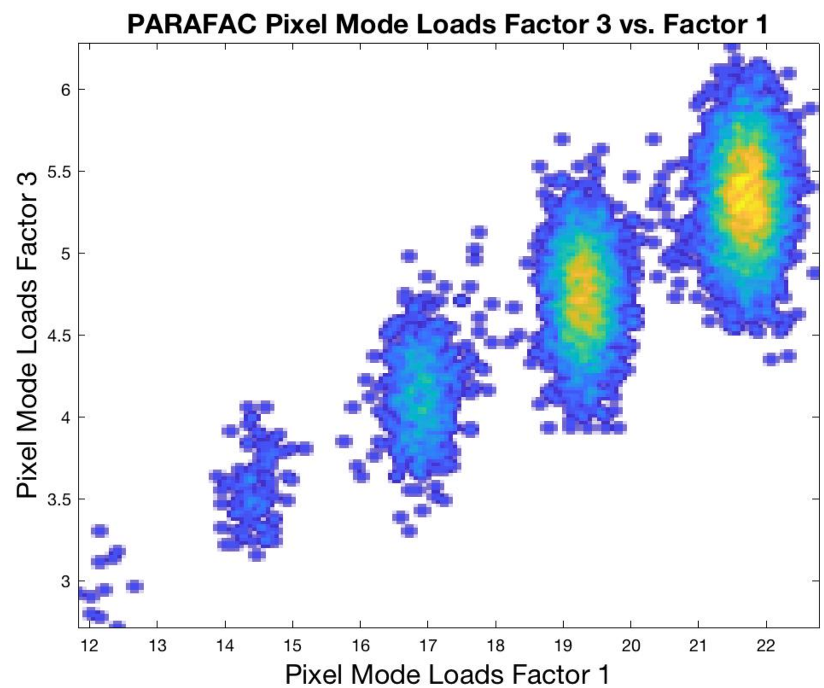






Pixel Factors Correlated!





- Variations in signal strength at each pixel persist as depth is profiled
- Seems to be 5 discrete signal strength levels
- Sampling problem??

EIGENVECTOR RESEARCH INCORPORATED

Software

- MATLAB®
- PLS_Toolbox
- MIA_Toolbox

EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- Many ways to look at data!
- "Color-by" especially useful on raw data or for adding info to other plots
- Factor based methods (PCA, MCR, PARAFAC, etc.) can condense data down to lower dimensions while retaining info
- Slicing and adding motion can reveal things previously un-noticed

EIGENVECTOR RESEARCH INCORPORATED