Synergy of Target and Anomaly Detection in Hyperspectral Images

Neal B. Gallagher

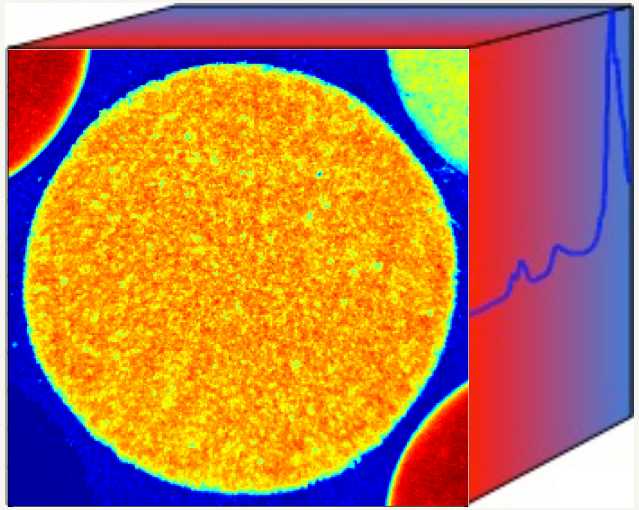
Eigenvector Research, Inc.

Booth 2652

Gallagher, Pittcon, Mar 9-12, 2015

EIGENVECTOR RESEARCH INCORPORATED

Hyperspectral Imaging



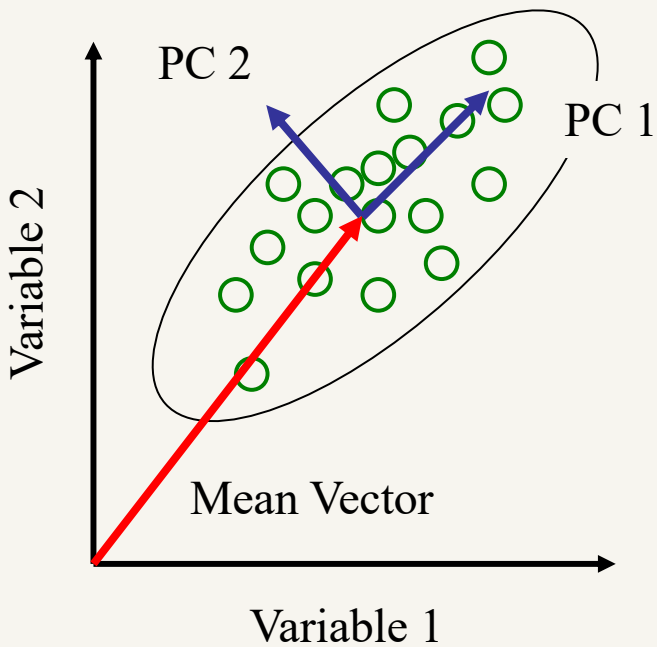
- Many pixels provide a good sampling for exploratory analysis, detection and classification
  - rapid and non-invasive
- Good for heterogeneous mixtures
  - although quantities to be detected may be low on a volume basis, signal in individual pixels can be dominated by an analyte of interest
  - wet chemistry methods can be hampered by dilution effects

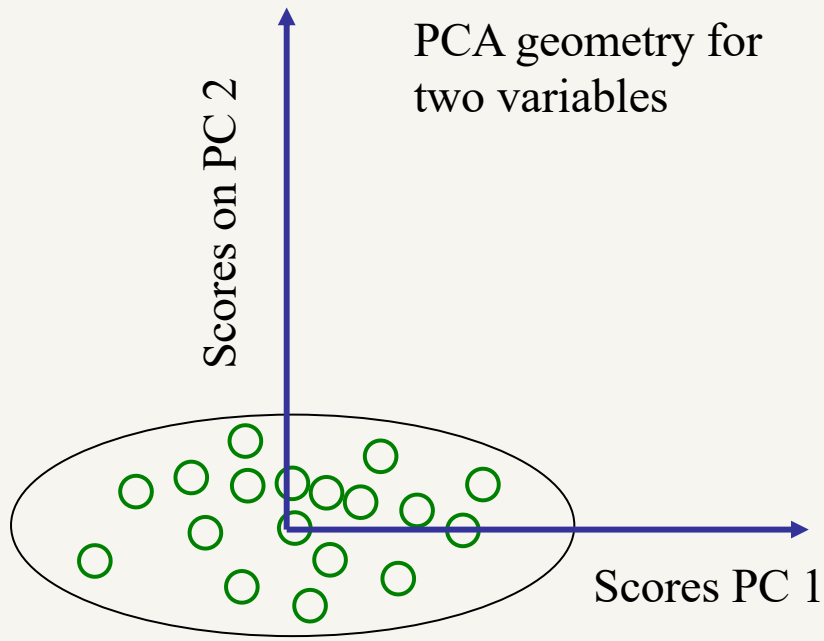
Typical Algorithms for Hyperspectral

- Exploratory Analysis and Anomaly detection
  - finds unusual pixels w/in an image in “any direction”
- Target detection
  - finds pixels w/ unusual signal in “specific directions” w/in an image
- Change detection
  - finds unusual pixels inter-image

Anomaly Detection

- PCA, others, …
  - Based on exploratory analysis algorithms





EIGENVECTOR RESEARCH INCORPORATED

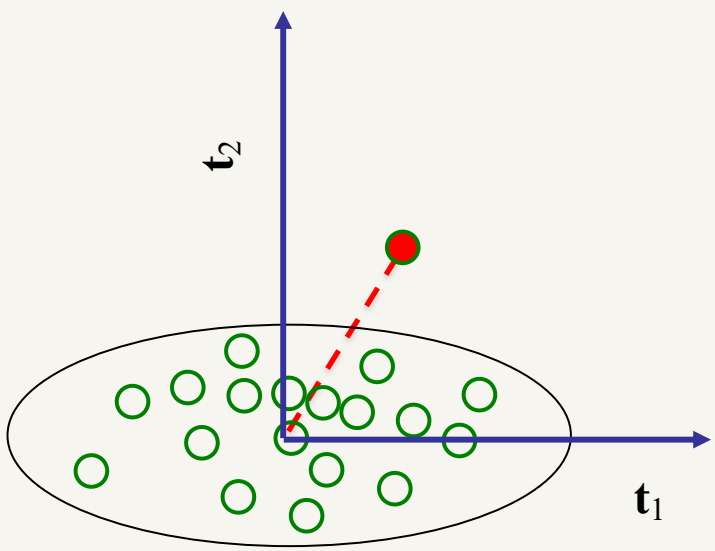
PCA Anomaly Detection

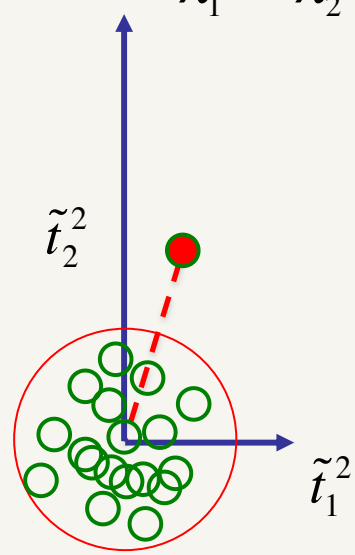
- Hotelling's T^2
  - Hotelling, H, (1931) "The generalization of Student's ratio," *Annals of Mathematical Statistics*," **2**(3), 360-378

$$T^2 = \mathbf{t}\left(\tfrac{1}{M-1}\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{t}^T$$

$$T^2 = \frac{t_1^2}{\lambda_1^2} + \frac{t_2^2}{\lambda_2^2} = \tilde{t}_1^2 + \tilde{t}_2^2$$

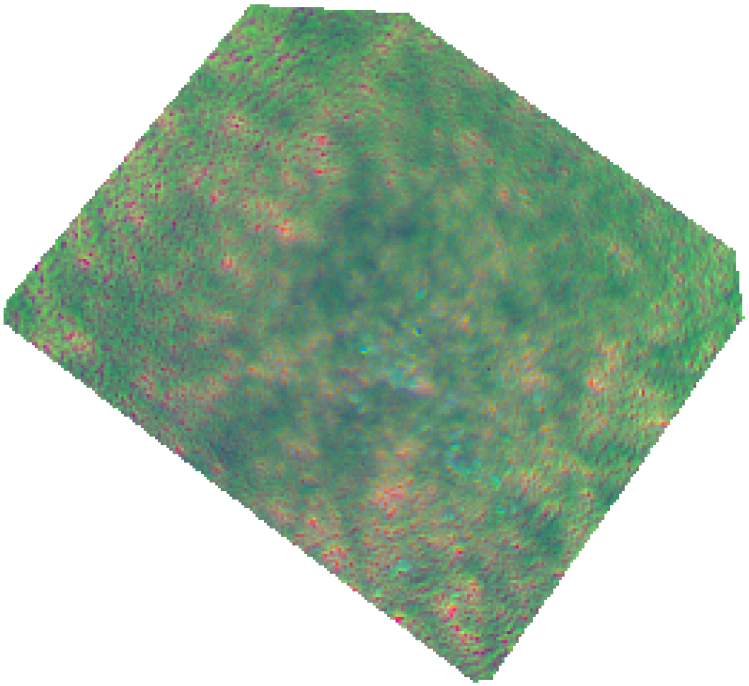
"whitening"





Anomaly Detection Example

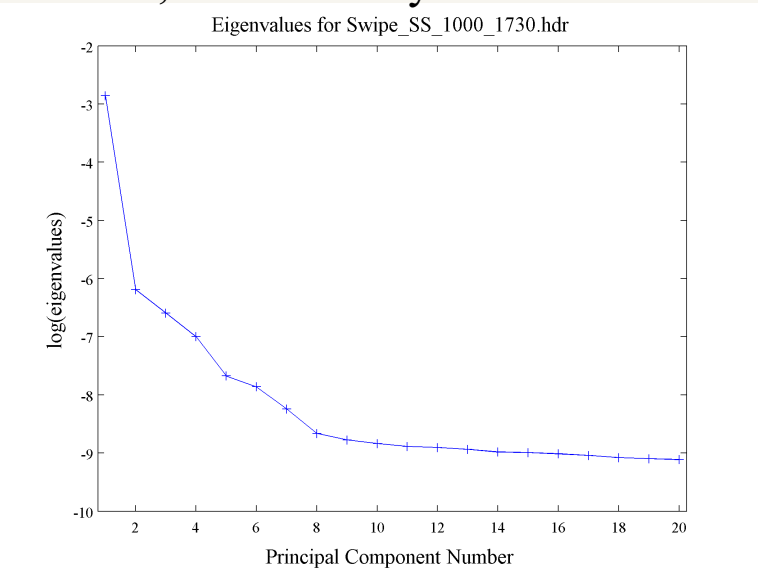
Image of Scores on PC 1 (97.73%), PC 2 (1.28%) & PC 3 (0.56%)



NIR reflectance image of a cellulosic swipe with ??? on it.

PCA scores 1,2,3 (RGB) for
-log10, mean-centered data.

a) is an anomaly apparent?
b) where is it?
c) if seen, can the analyte be identified?

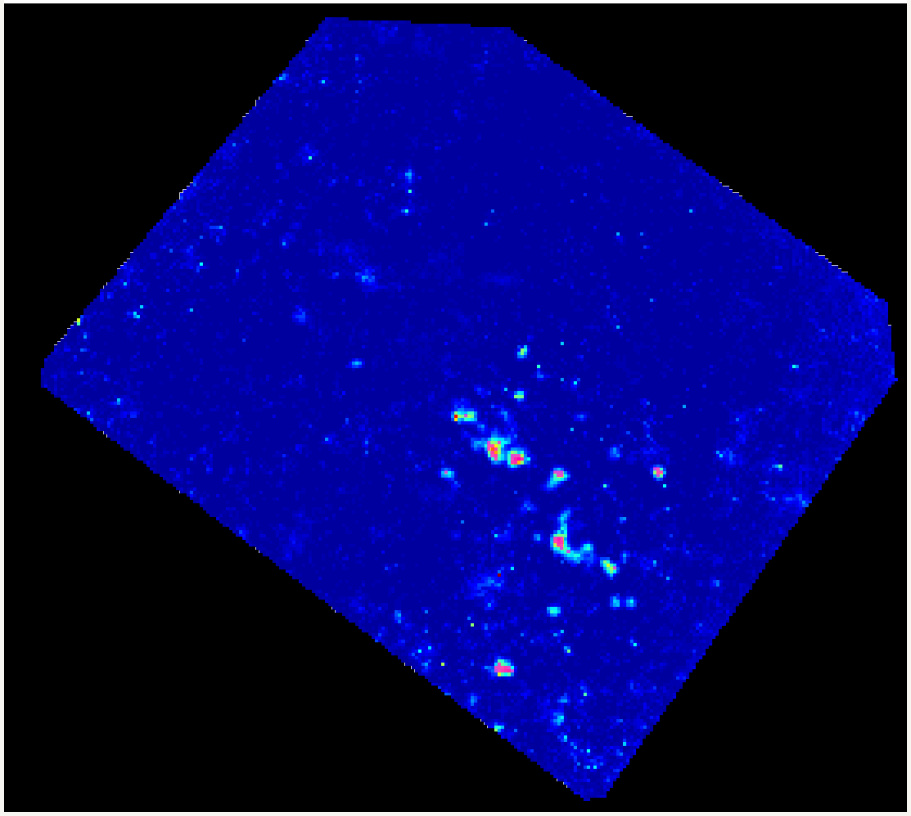


data set: courtesy OPOTEK, Inc., Carlsbad, CA

EIGENVECTOR RESEARCH INCORPORATED

T^2 Anomaly Detection

Image of Hotelling T^2 (99.83%)



weak anomaly found on PCs 4 to 6

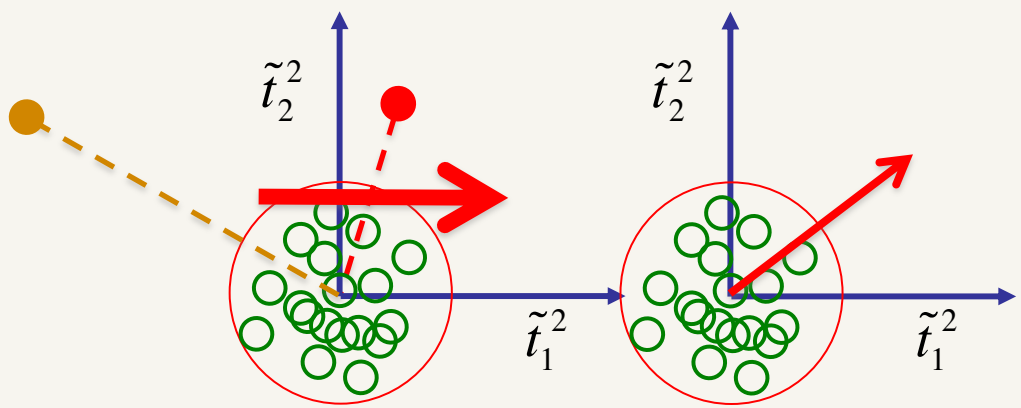
X-block: Swipe_SS_1000_1730.hdr 35208 by 731
Preprocessing: -log10, Mean Center
Num. PCs: 7

Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	1.03e+01	97.73	97.73
2	1.35e-01	1.28	99.01
3	5.84e-02	0.56	99.57
4	1.86e-02	0.18	99.75
5	4.28e-03	0.04	99.79
6	3.17e-03	0.03	99.82
7	1.78e-03	0.02	99.83

Anomaly Detection Summary

- Finds unusual pixels in *any* direction
- Doesn't tell *what* the anomaly is due to
- Is it possible to utilize what is known about target analyte to choose *a specific* direction?



Target Detection

- Finds pixels w/ unusual signal in a *specific direction* w/in an image
- Provides information on *what* an anomaly is
- Algorithms:
  - Generalized least squares (GLS), spectral angle mapper, window target factor analysis
- For target detection it is nice to know how the signal manifests in general
  - …explore the signal

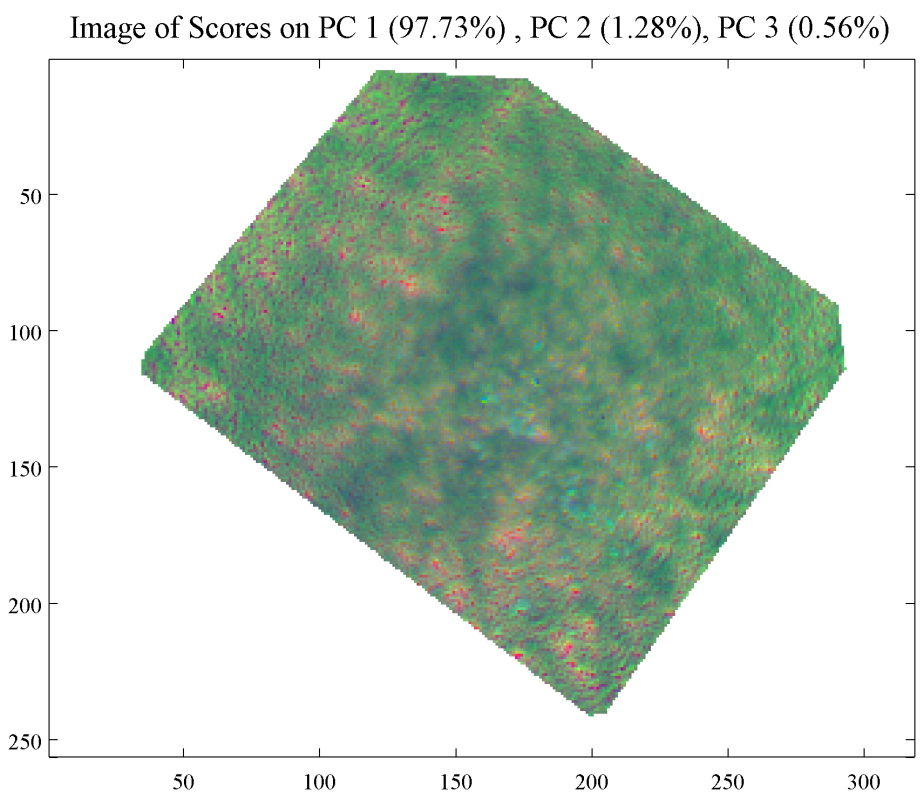
Target Detection

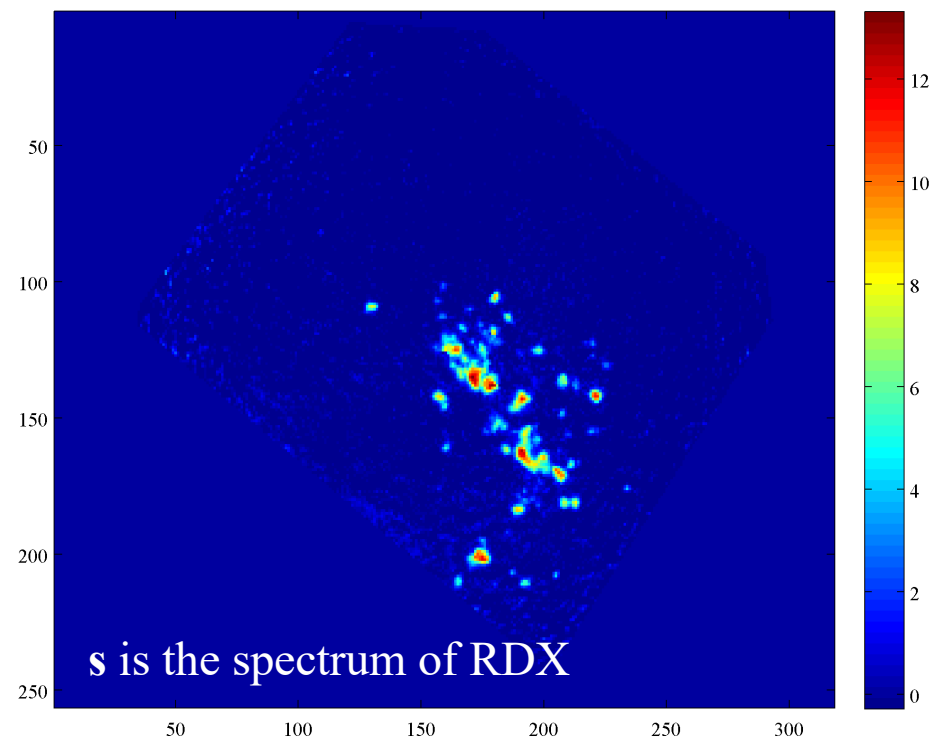
$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

$$\mathbf{T} = \mathbf{X}\mathbf{P}$$

$$\mathbf{X} = \mathbf{c}\mathbf{s}^T + \mathbf{E}$$

$$\hat{\mathbf{c}} = \mathbf{X}\boldsymbol{\Sigma}_c^{-1}\mathbf{s}\left(\mathbf{s}^T\boldsymbol{\Sigma}_c^{-1}\mathbf{s}\right)^{-1}$$



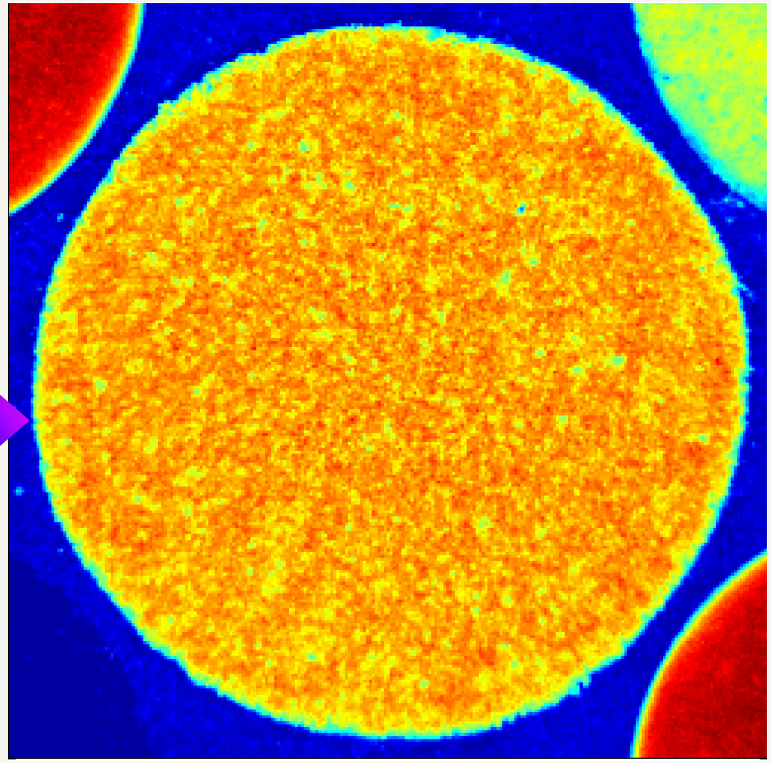


data set: courtesy OPOTEK, Inc., Carlsbad, CA

Model System Studied

- ppm levels of melamine in wheat gluten

NIR reflectance hyperspectral image



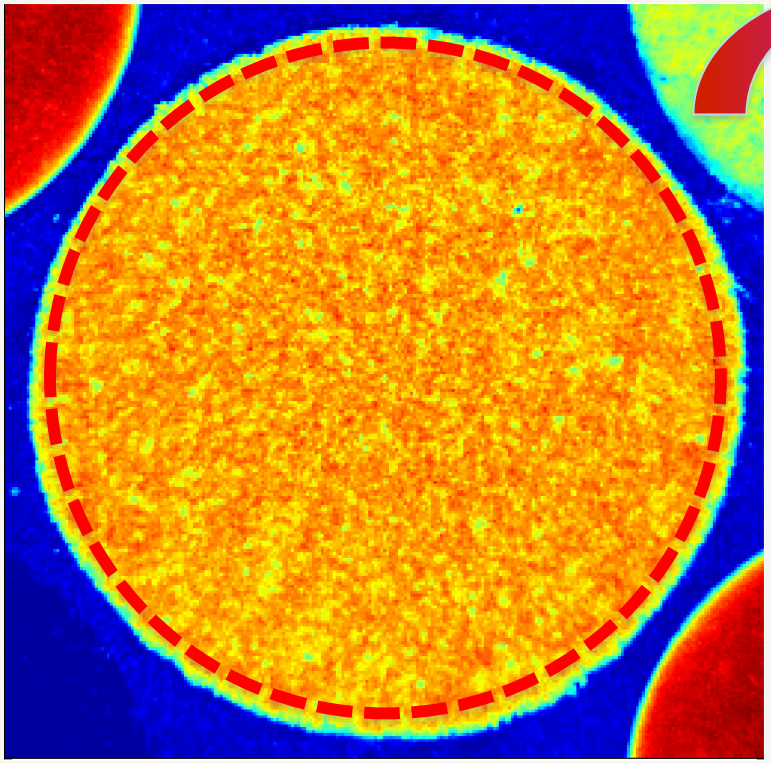
EIGENVECTOR RESEARCH INCORPORATED

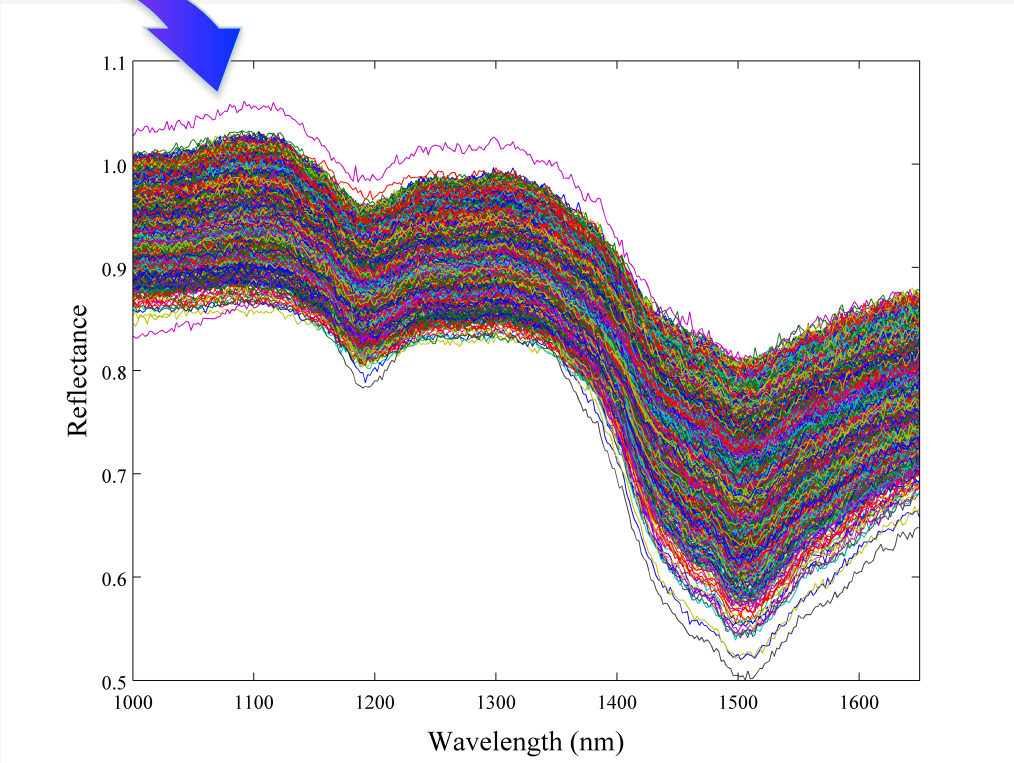
Powdered Raw Materials

- Challenges
  - scattering and particle size distributions changes sample-to-sample
  - same material from a wide variety of sources
- Calibration for typical ILS models difficult
  - unlikely to acquire a calibration data set that spans all the sample variation expected to be seen
    - and if you do, the net analyte signal suffers
  - unlikely to use one ILS model from a single material for multiple raw materials

Unadulterated Wheat Gluten

Signal from the unadulterated wheat gluten is highly variable and is much stronger than the adulterant.





Unadulterated wheat gluten

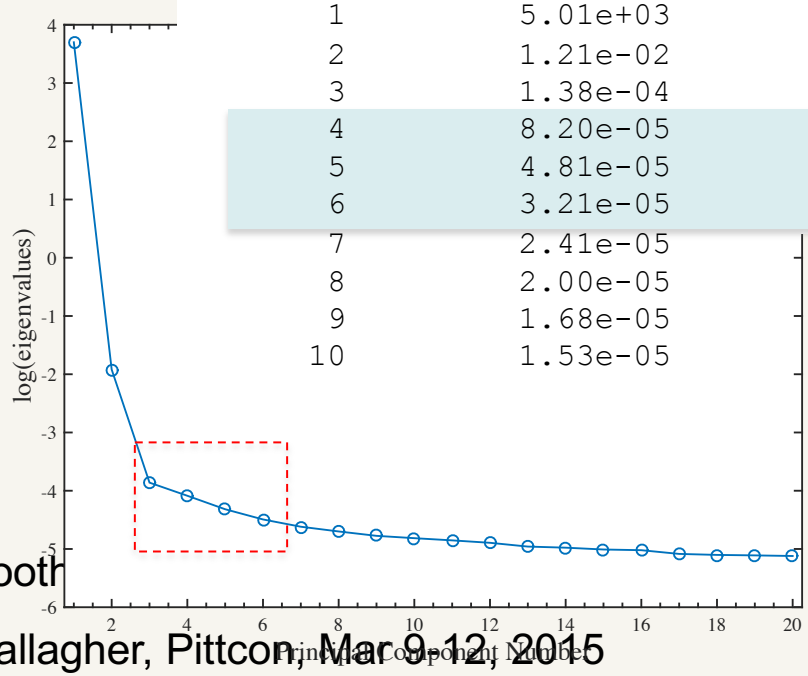
EIGENVECTOR RESEARCH INCORPORATED

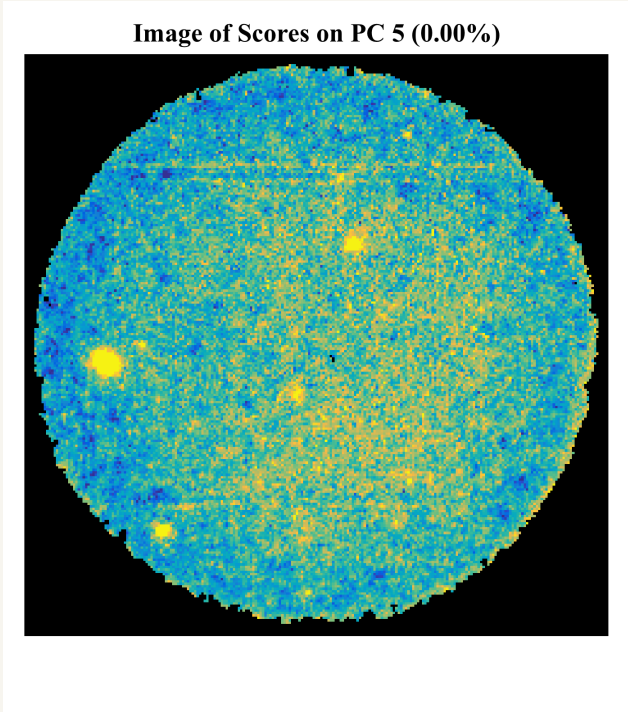
PCA of Melamine in Wheat Gluten

Eigenvalues for 200 ppm melamine in wheat gluten
-log10

Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC
1	5.01e+03	100.00
2	1.21e-02	0.00
3	1.38e-04	0.00
4	8.20e-05	0.00
5	4.81e-05	0.00
6	3.21e-05	0.00
7	2.41e-05	0.00
8	2.00e-05	0.00
9	1.68e-05	0.00
10	1.53e-05	0.00

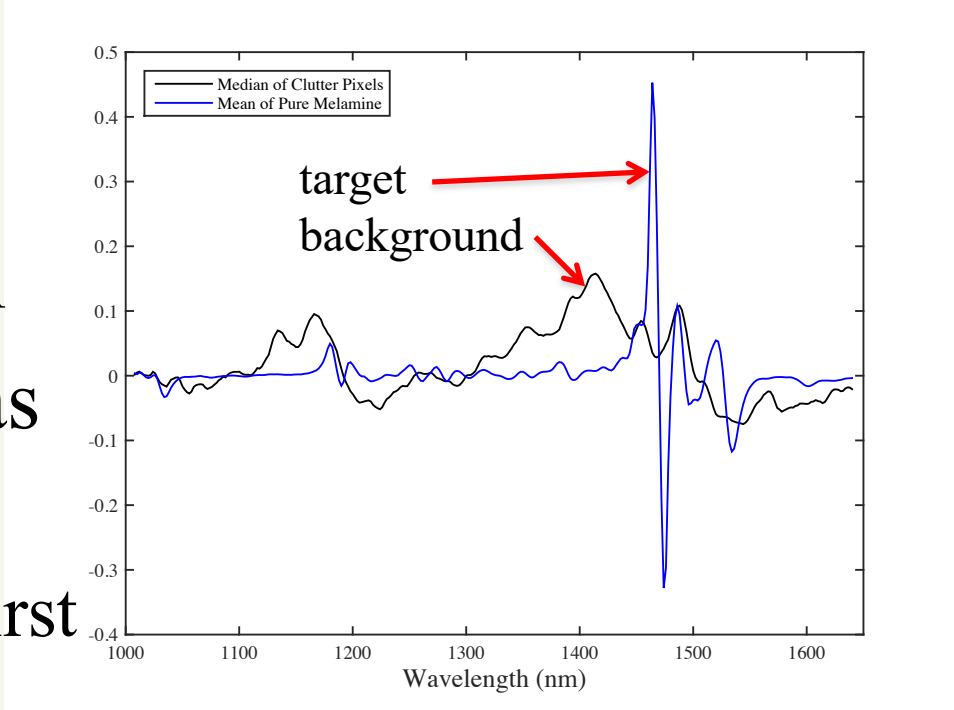


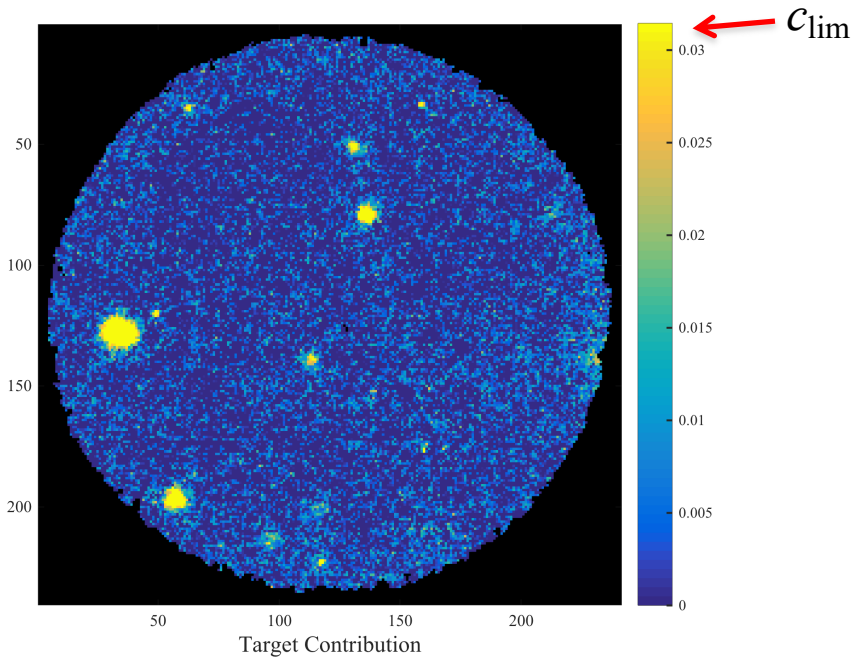


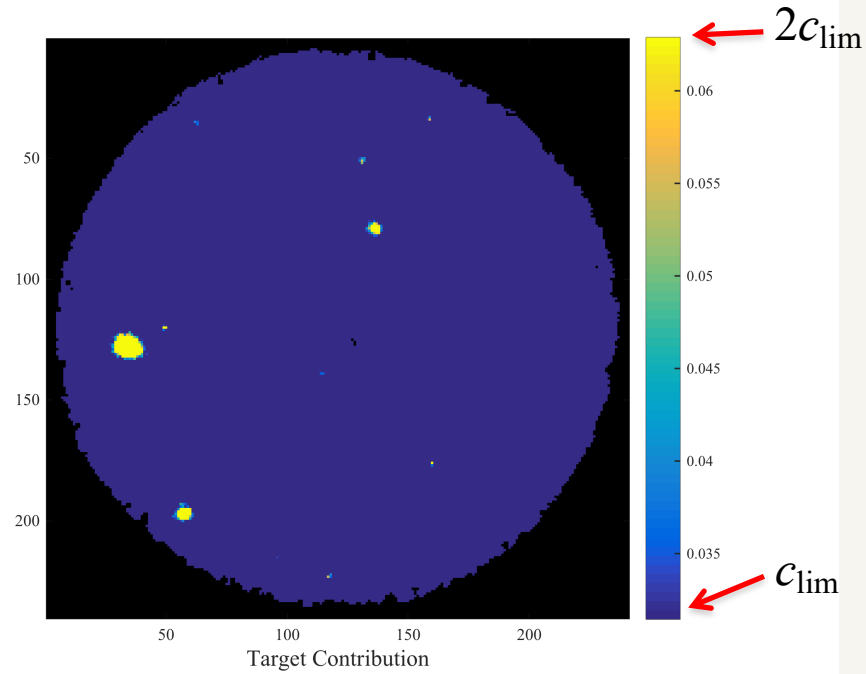
EIGENVECTOR RESEARCH INCORPORATED

Target Approach

- 2nd derivative preprocessing
  - eliminates much of the irrelevant clutter signal
- Use library spectrum as the target
  - libraries can be good first approximations





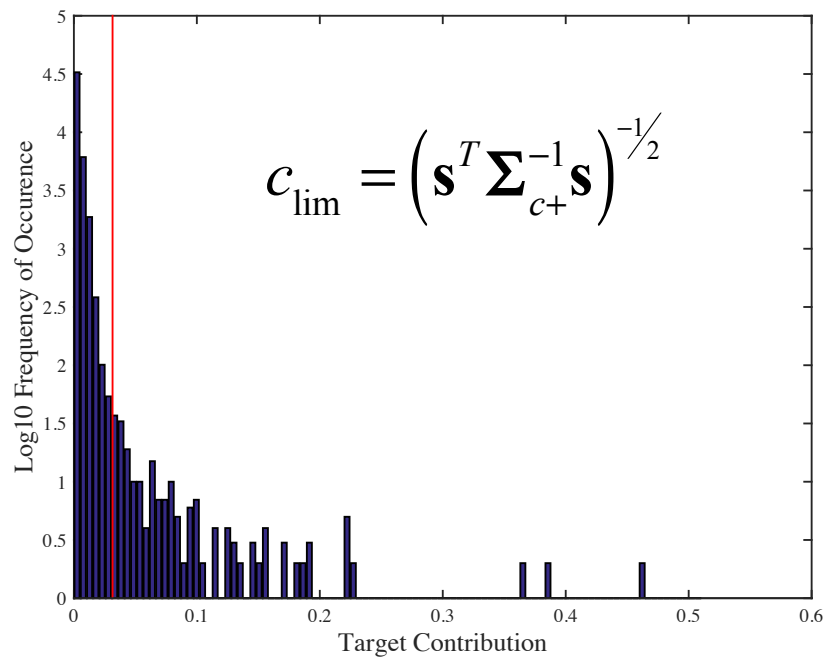


target contributions

visualization saturates the image signal to

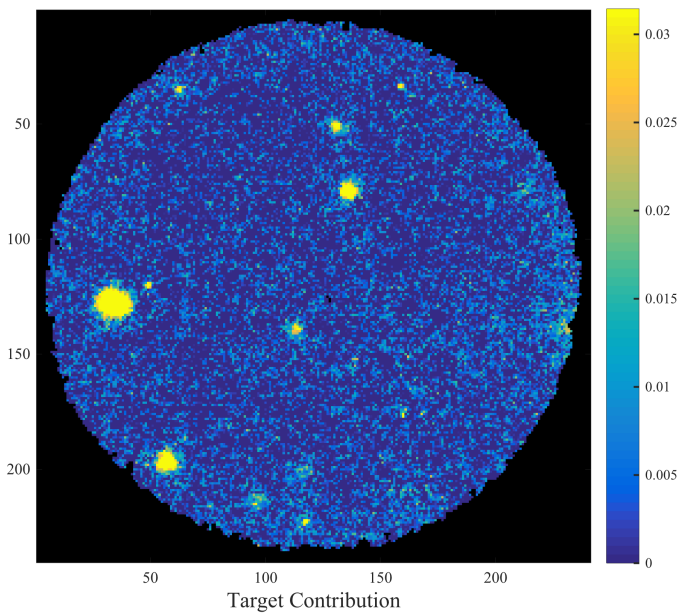
$$c_{\mathrm{lim}} < c < 2c_{\mathrm{lim}}$$

whitening based on all the pixels



Whitened PCA Anomaly Detection

GLS target detection results.



- Define pixels with $c > c_{lim}$ as $\mathbf{X}_s$ and whiten using all other pixels.
- Known target pixels are "signal" and others are "clutter". We are now supervising the PCA.

$$\boldsymbol{\Sigma}_c^{-1/2}\mathbf{X}_s^T\mathbf{X}_s\boldsymbol{\Sigma}_c^{-1/2}\mathbf{p} = \lambda\mathbf{p}$$

$$\boldsymbol{\Sigma}_c = \frac{1}{M_c}\mathbf{X}_c^T\mathbf{X}_c$$

- If $\mathbf{X}_s$ is a single target pixel, $\mathbf{x}_s$, the approach reduces to GLS target detection.
- The pixels selected in the image are generally expected to be more relevant as a target than library spectra.
- Whitening based on non-target pixels

Booth 2652

Gallagher, Pittcon, Mar 9-12, 2015

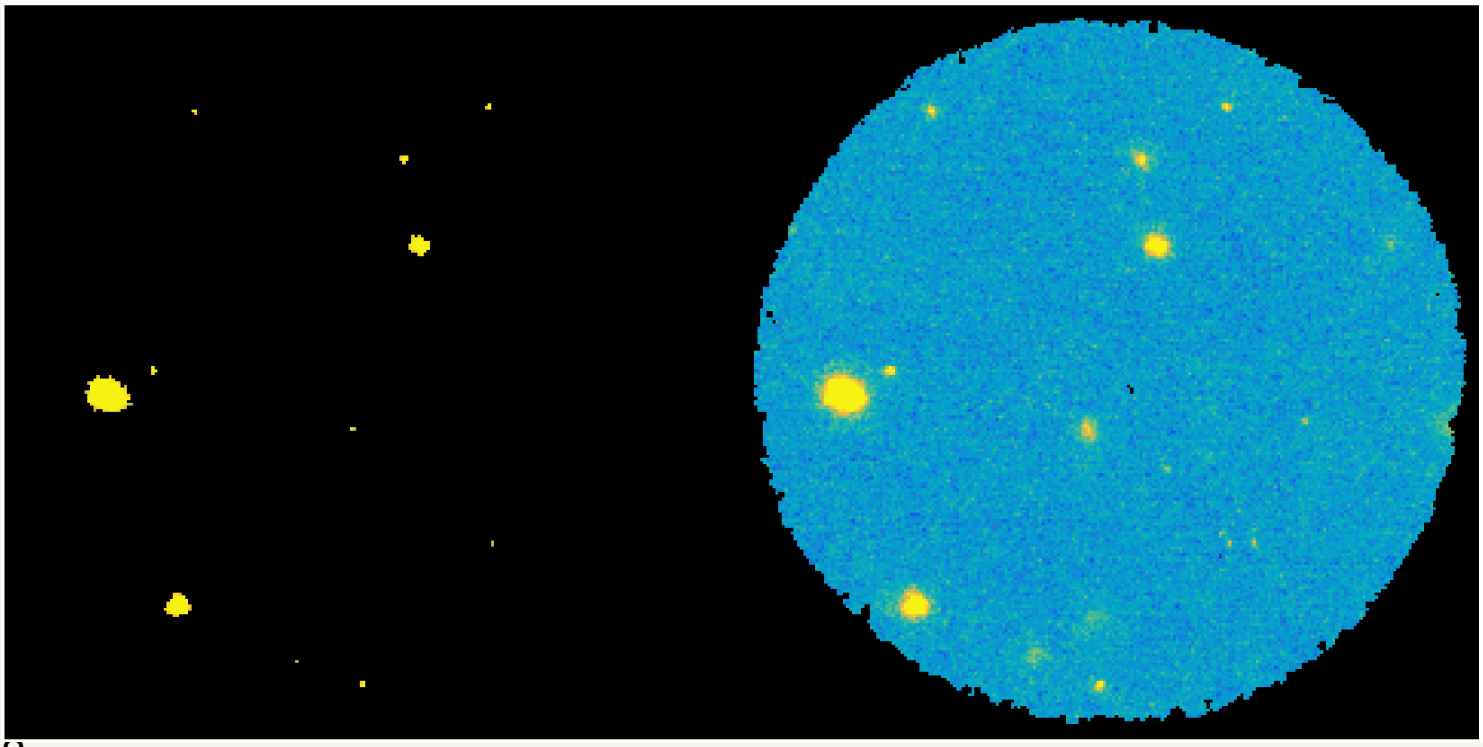
EIGENVECTOR RESEARCH INCORPORATED

Calibrate on Whitened Target Pixels

Pixels identified in target detection are used as "image relevant target signal."

Off-target pixels are used for whitening (= "clutter signal")

Image of Scores on PC 1 (94.26%)



Targeted Anomaly Detection

- Target detection based on a library spectrum identified candidate target-rich pixels
- Target-rich pixels replace the library spectrum as target signal, and
- Off-target-rich pixels are clutter and are used to de-weight clutter signal
  - clutter = all signal not of interest / non-target signal
- Whitened PCA is "targeted anomaly detection"

Image of Scores on PC 1 (94.26%)

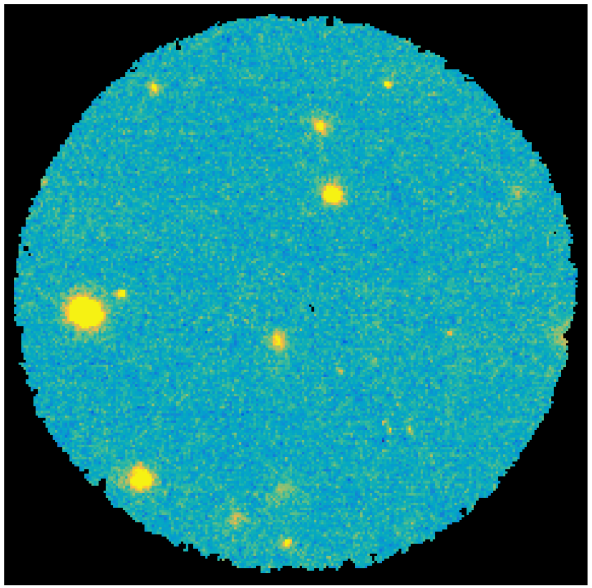
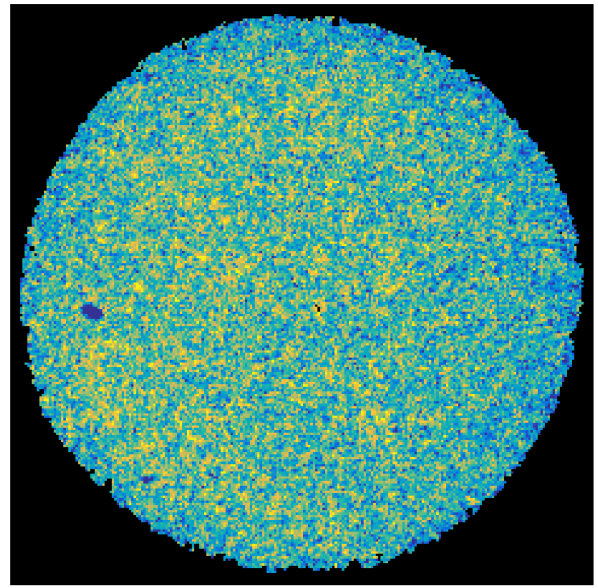
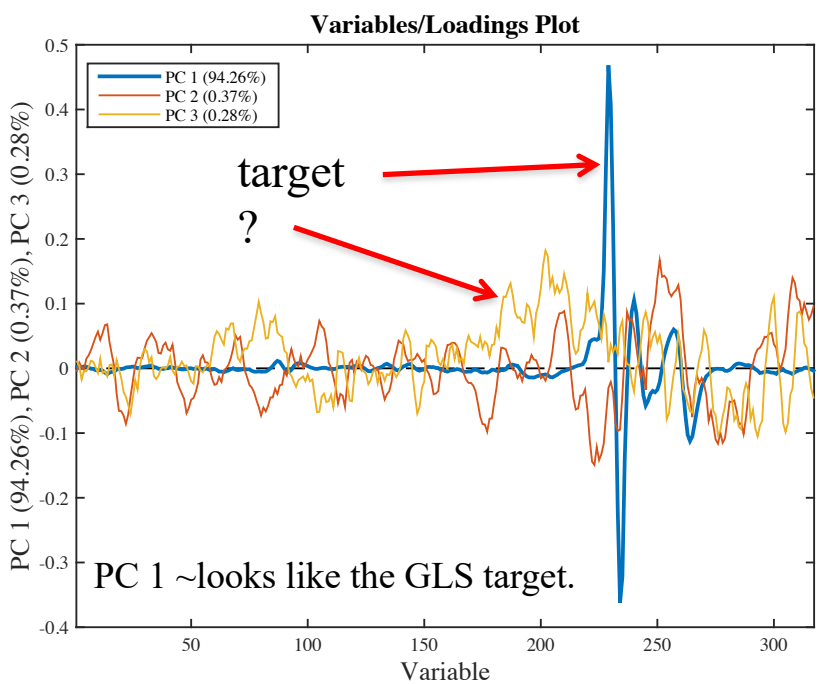


Image of Scores on PC 3 (0.28%)





1st derivative

center to mean of clutter

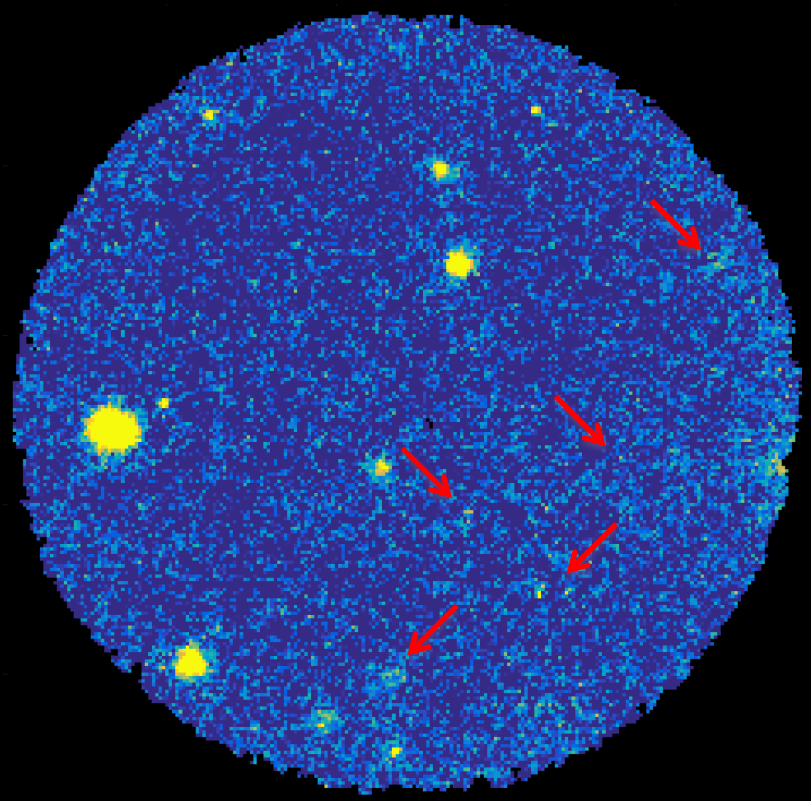
autocontrast the scores

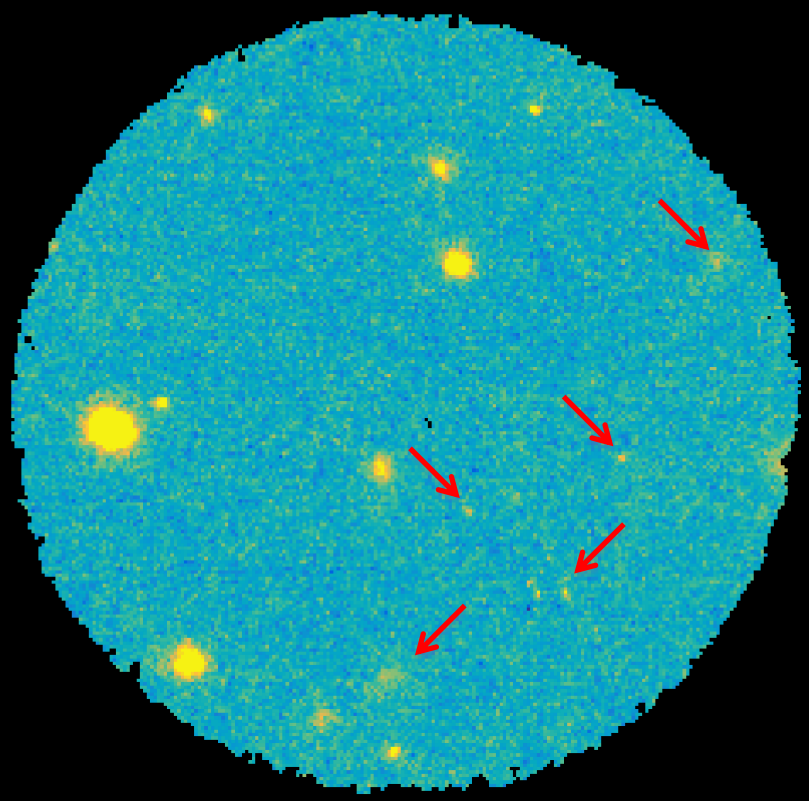
Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC
1	2.27e+02	94.26
2	8.93e-01	0.37
3	6.70e-01	0.28

EIGENVECTOR RESEARCH INCORPORATED

Compare GLS & WPCA





EIGENVECTOR RESEARCH INCORPORATED

Image of Scores on PC 1 (71.80%)

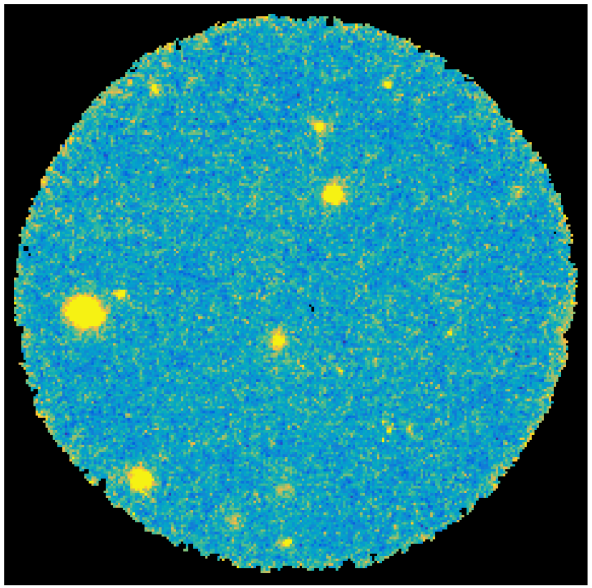


Image of Scores on PC 2 (3.18%)

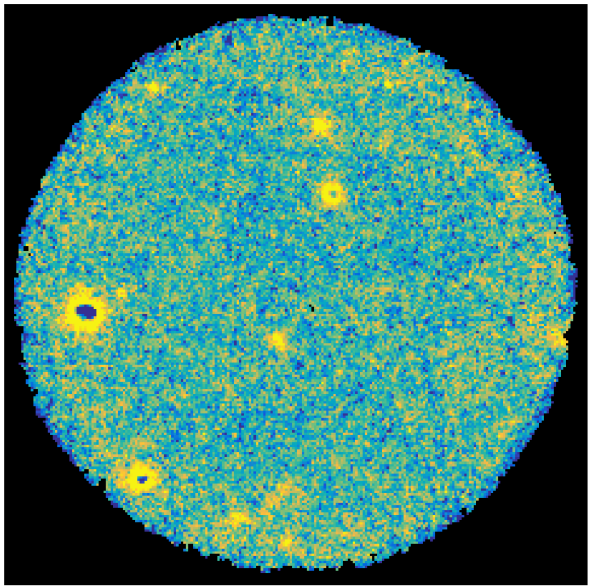
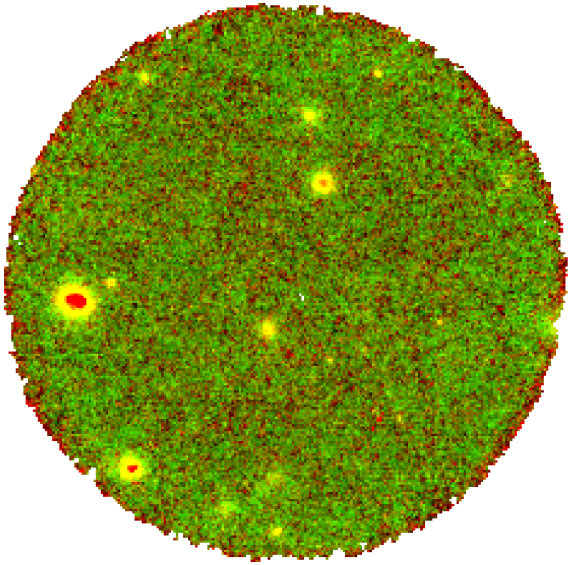
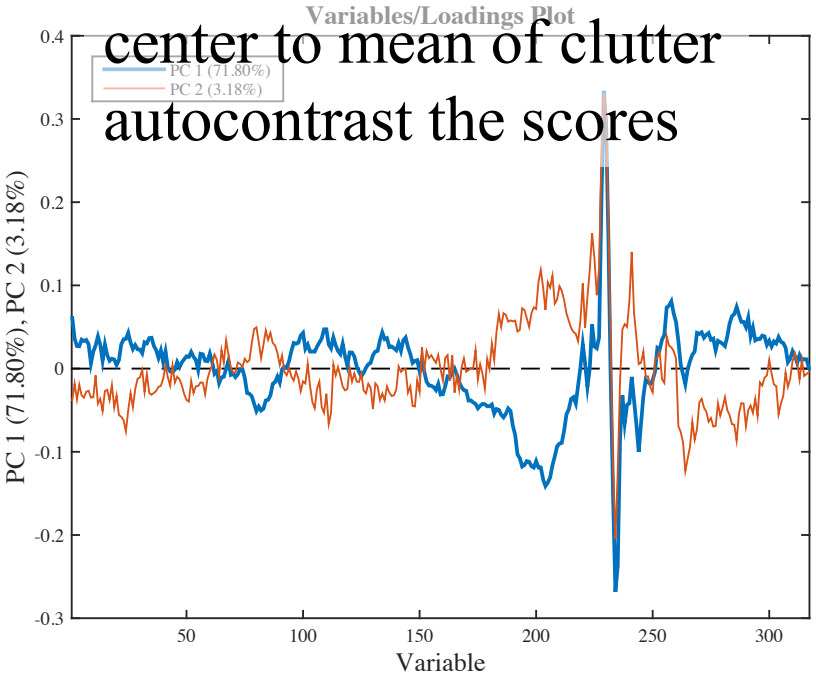


Image of Scores on PC 1 (71.80%) & Scores on PC 2 (3.18%)



1st derivative, SNV
center to mean of clutter
autocontrast the scores



Conclusions

- Anomaly and Target detection can be used synergistically to find signal of interest
  - signal that may be missed when used independently
- The math behind many methods is similar
  - e.g., GLS and WPCA
  - as shown WPCA can be considered "targeted anomaly detection"
- Differences are in how "signal" and "clutter" are defined
  - Can it be improved further with additional iterations? Automated?