Are O-PLS Models Really More Interpretable?

Barry M. Wise and Jeremy M. Shaver
Eigenvector Research, Inc., Wenatchee, USA

O-PLS

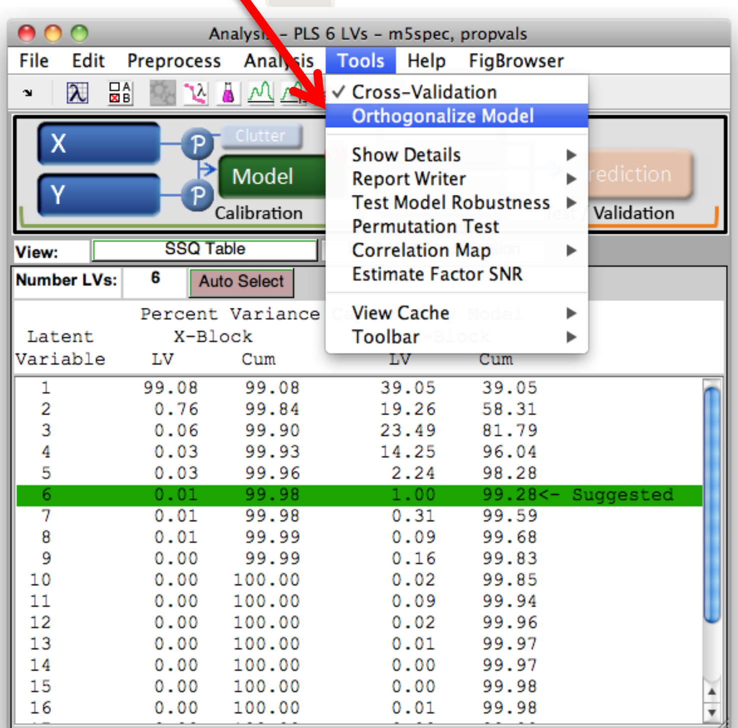
- Originally formulated as sequential algorithm (NIPALS based)
- Since shown to be obtainable from post-processing conventional PLS model
- Does not improve prediction
- Claim is that model is more interpretable

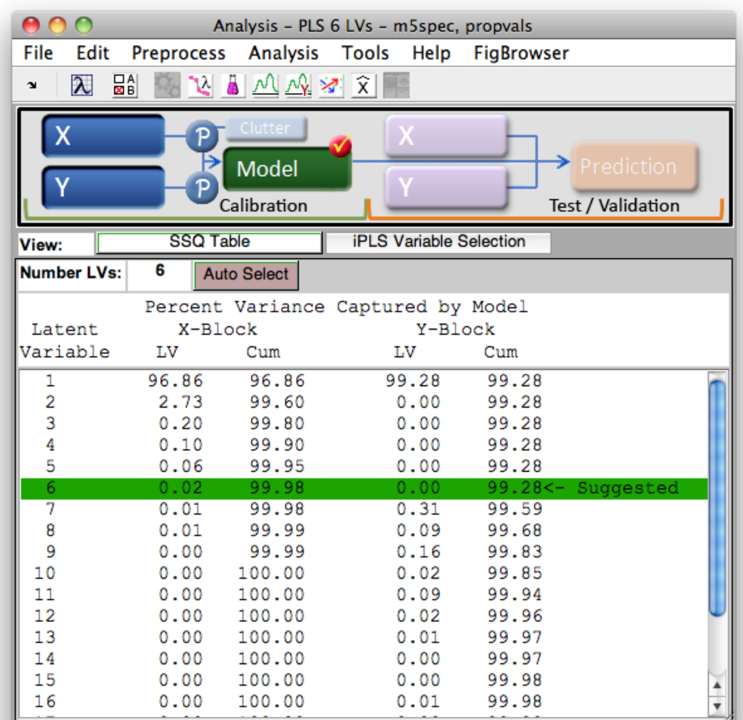
E.K. Kemsley and H.S. Tapp, "OPLS filtered data can be obtained directly from non-orthogonalized PLS1," *J. Chemo*, **23**, 263-264, 2009

R. Ergon, "PLS post-processing by similarity transformation (PLS+ST): a simple alternative to OPLS," *J. Chemo*, **19**, 1-4, 2005

J. Trygg and S. Wold, "Orthogonal Projections to Latent Structures (O-PLS)," *J. Chemo*, **16**, 119-128, 2002.

Orthogonalize Model





EIGENVECTOR RESEARCH INCORPORATED

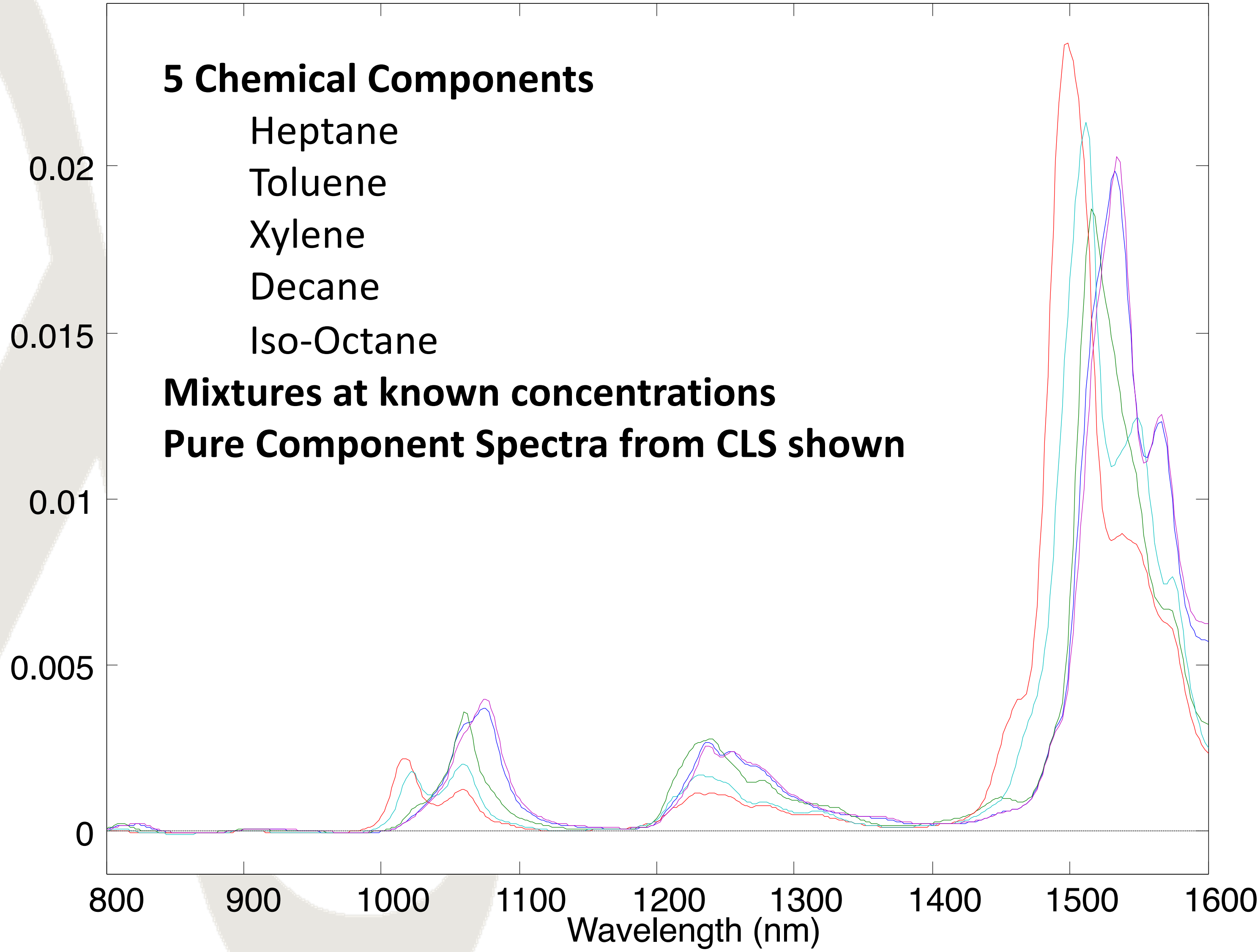
Questions:

- What do we need to be aware of when interpreting O-PLS recovered components?
- What kinds of sensitivities does O-PLS have to noise, rotational ambiguity, and correlated concentrations?

Method: Use well-characterized and/or carefully constructed simple systems to study OPLS

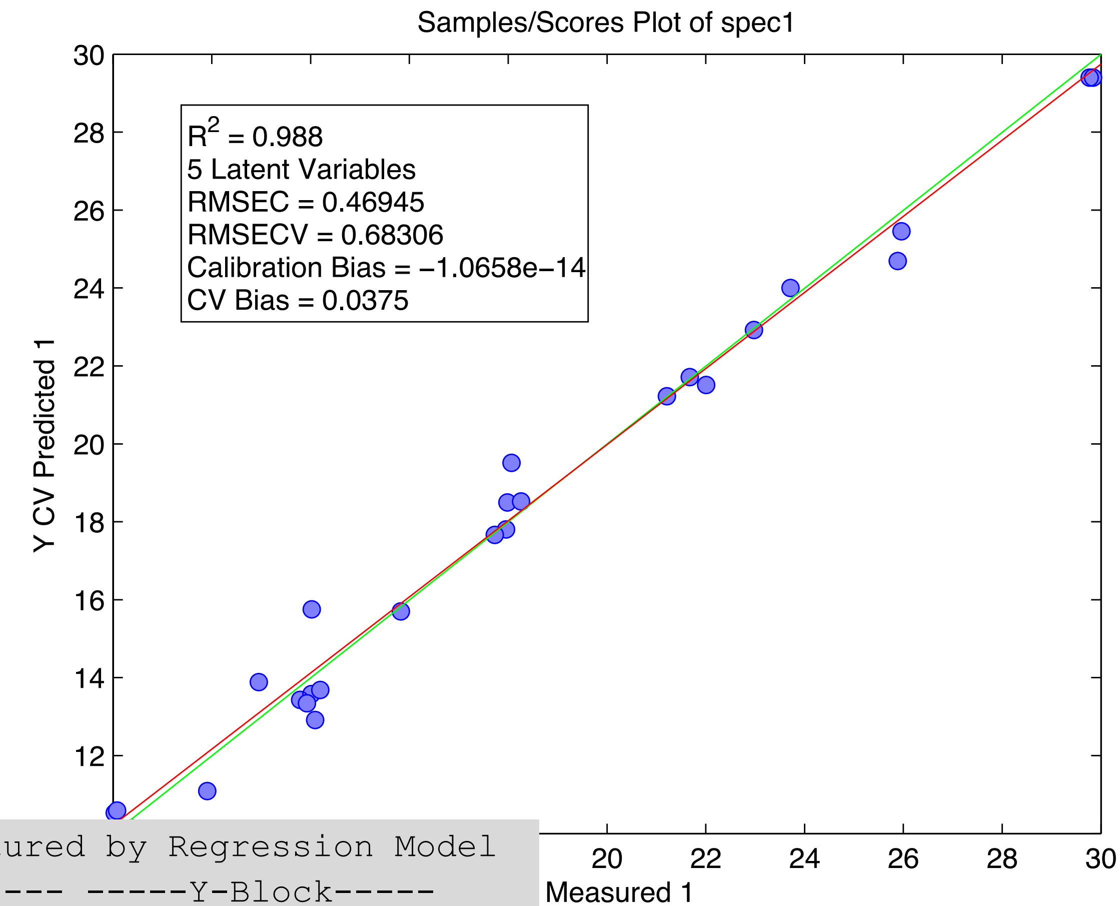
EIGENVECTOR RESEARCH INCORPORATED

NIR of Pseudo-gasoline Samples



EIGENVECTOR RESEARCH INCORPORATED

PLS Model of Heptane

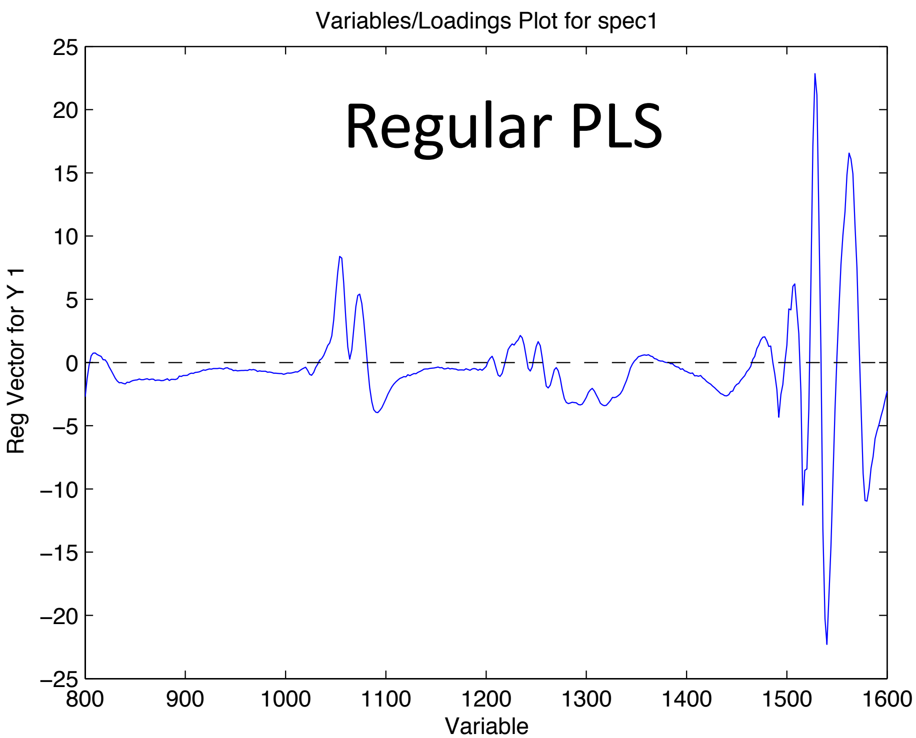


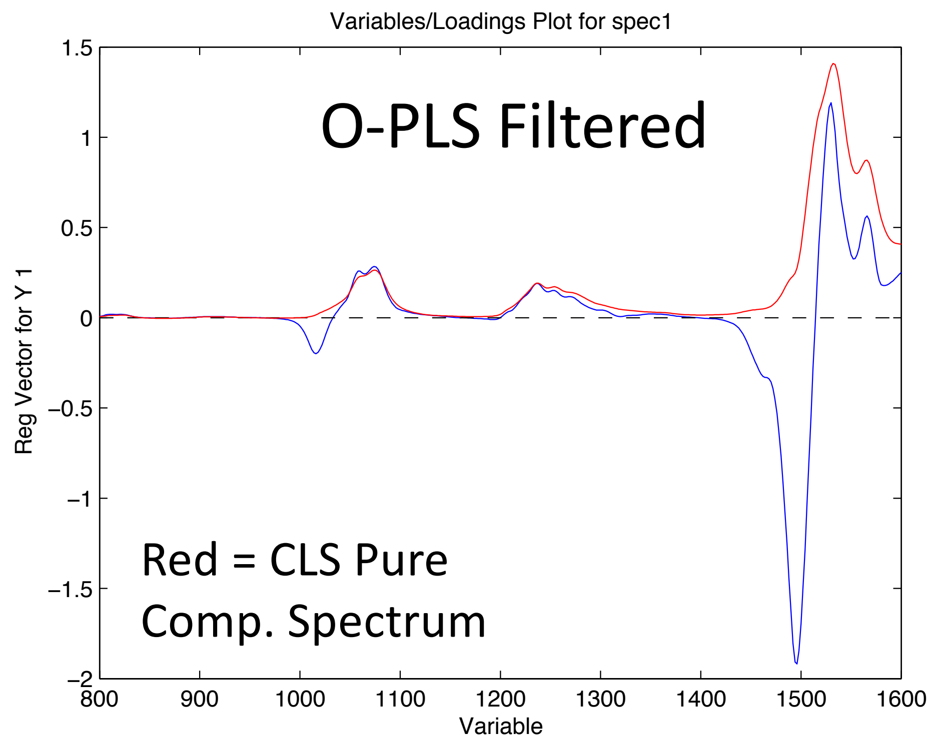
Percent Variance Captured by Regression Model

	-----X-Block-----		-----Y-Block-----	
Comp	This	Total	This	Total
1	91.17	91.17	8.36	8.36
2	7.40	98.57	7.19	15.55
3	0.93	99.50	32.81	48.36
4	0.46	99.96	26.18	74.54
5	0.02	99.98	24.90	99.44

EIGENVECTOR RESEARCH INCORPORATED

Regular PLS and O-PLS Filtered Regression Vectors

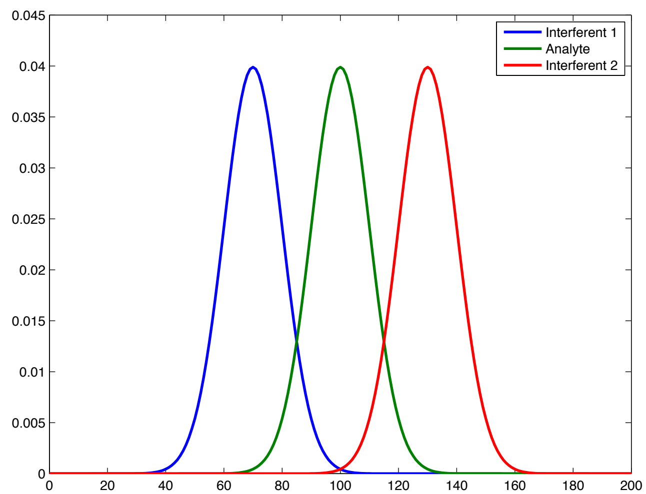




EIGENVECTOR RESEARCH INCORPORATED

Simple System Example

- Synthetic example of three constituents
- Evenly spaced Gaussian peaks, analyte in middle
- Vary correlation between analyte and interferents

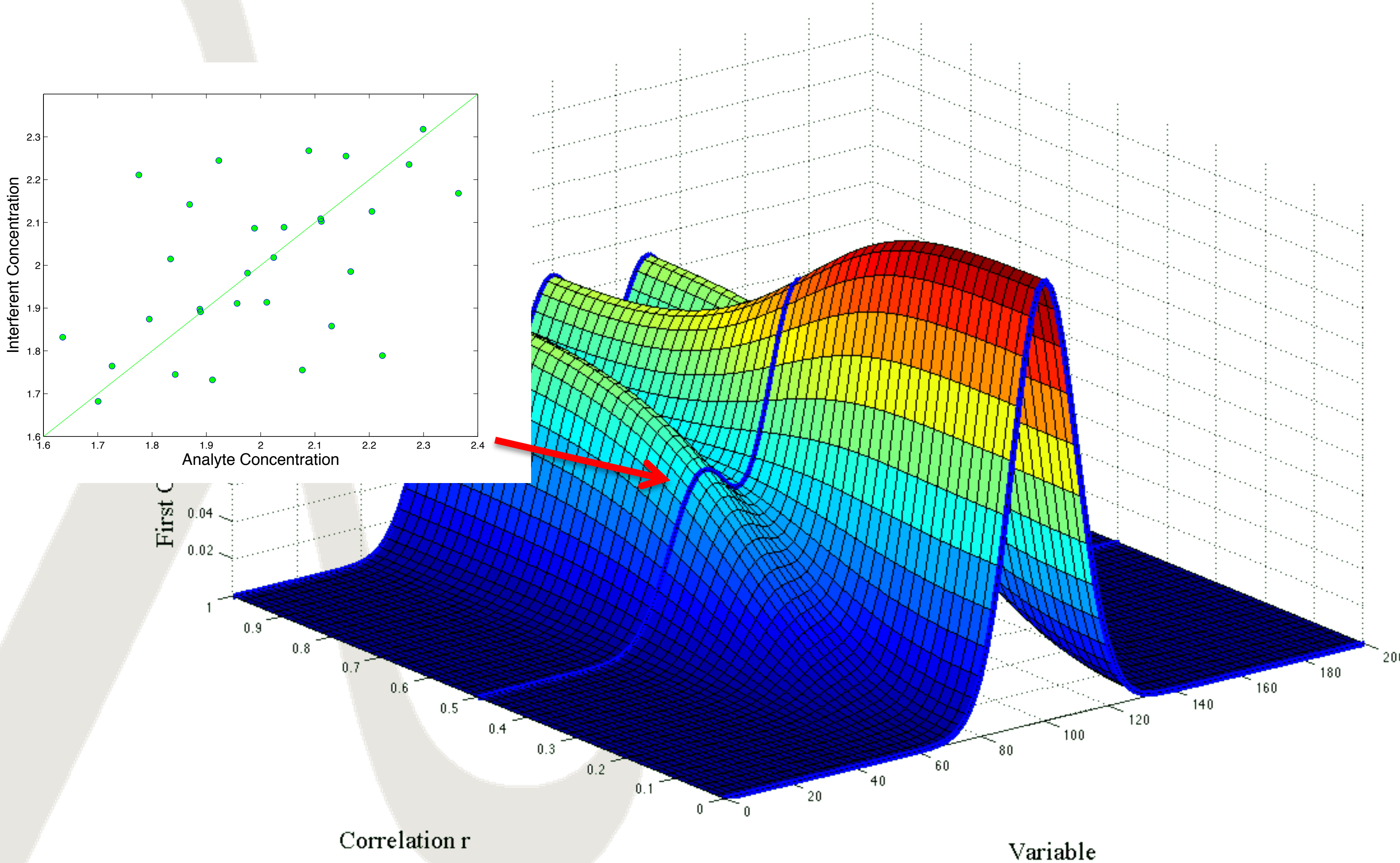


EIGENVECTOR RESEARCH INCORPORATED

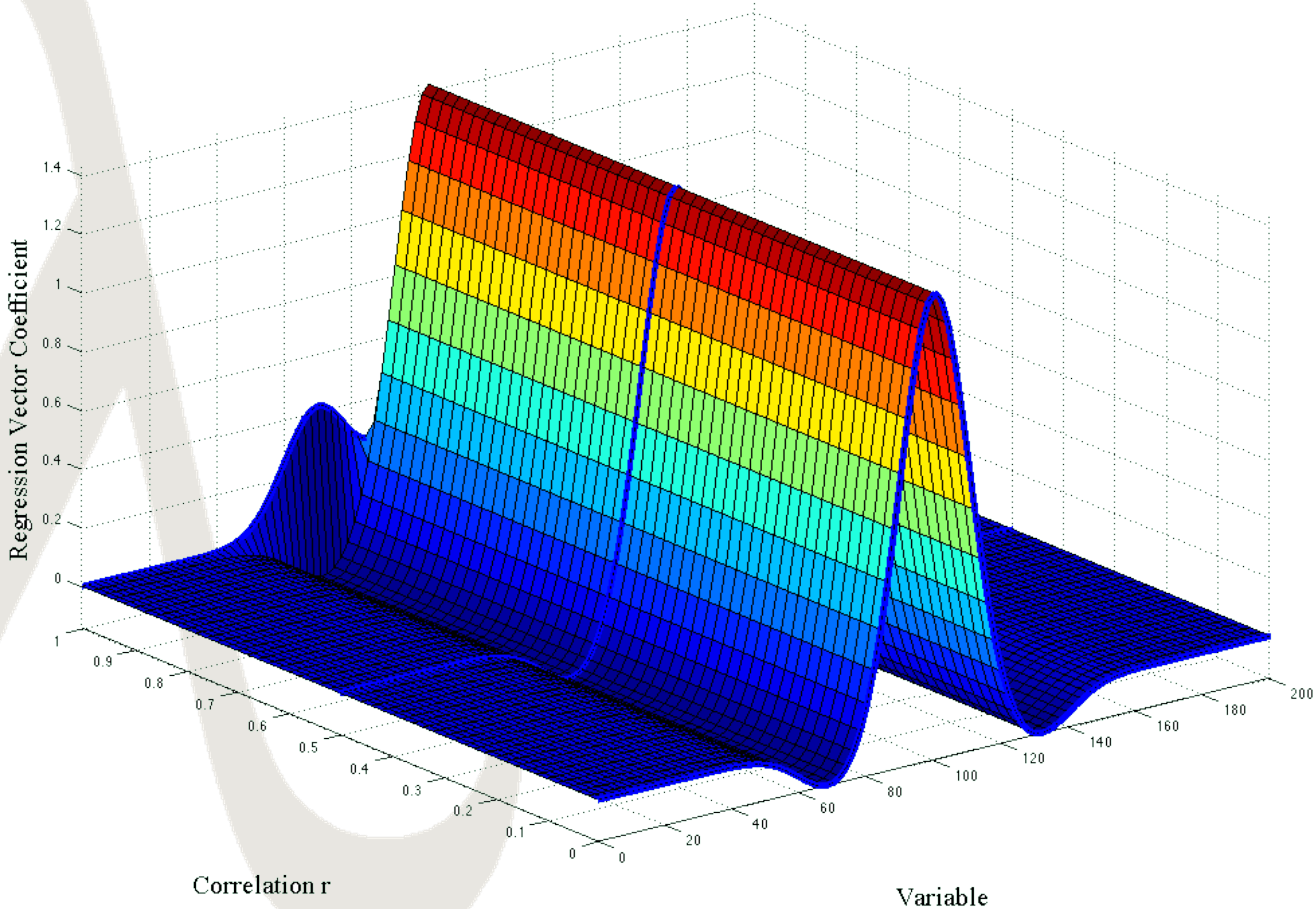
Gaussian Peaks Scenarios

- Start with orthogonal concentrations
  - 1) Go from orthogonal to positively correlated concentrations
  - 2) One interferent positively correlated, one negatively correlations

Scenario 1-First O-PLS Loading



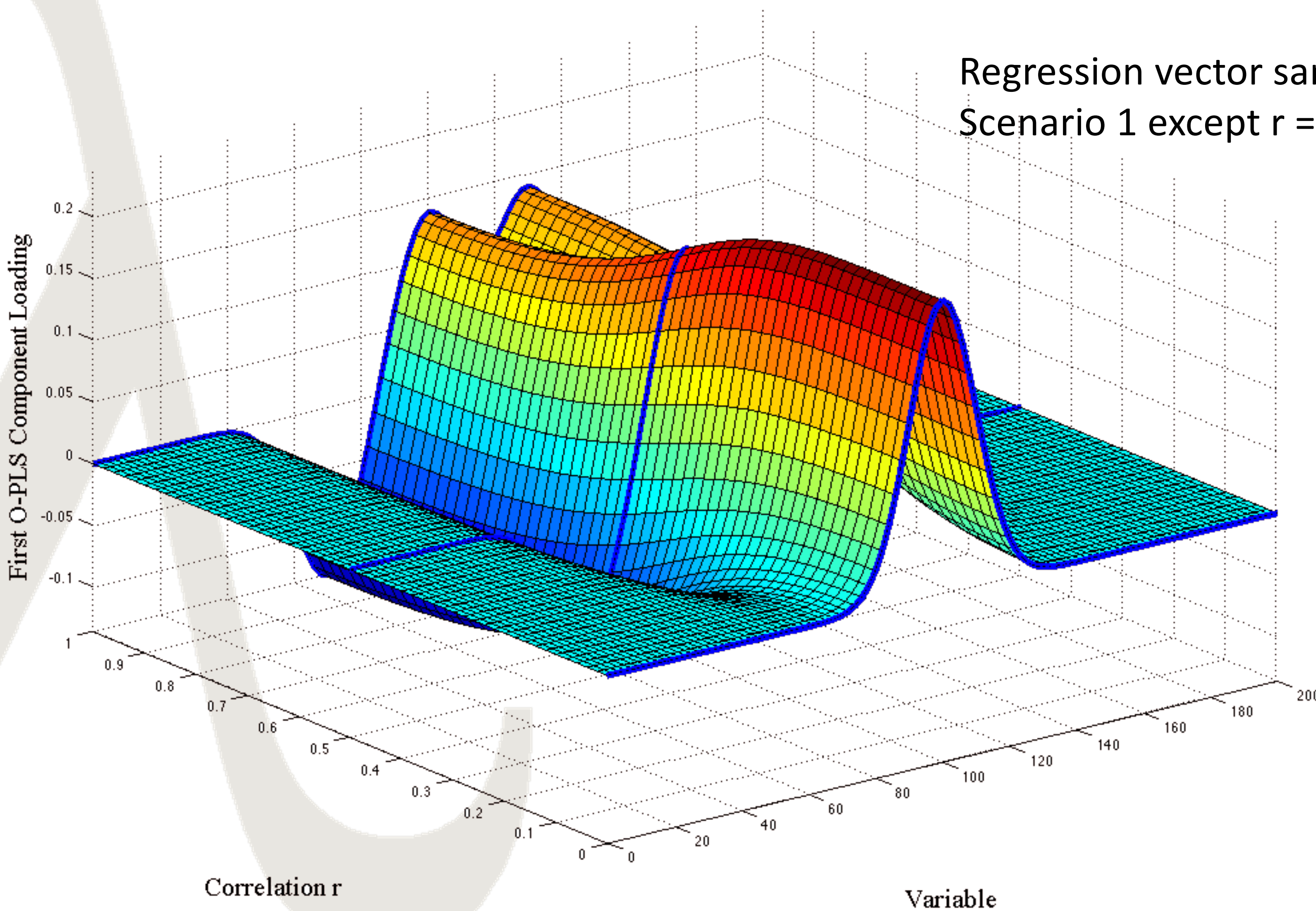
Scenario 1-Regression Vector



EIGENVECTOR RESEARCH INCORPORATED

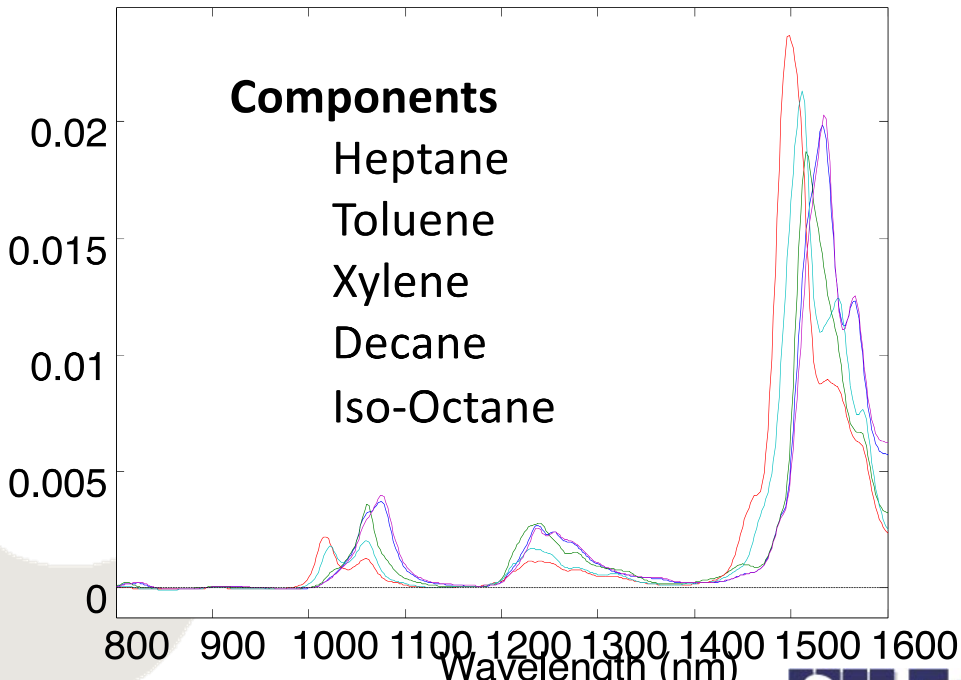
Scenario 2-First O-PLS Loading

Regression vector same as
Scenario 1 except r = 1 case



Pseudo-gasoline Example

- 5 component mixture measured by NIR
- Solve for pure components via CLS
- Use pure spectra to create synthetic scenarios

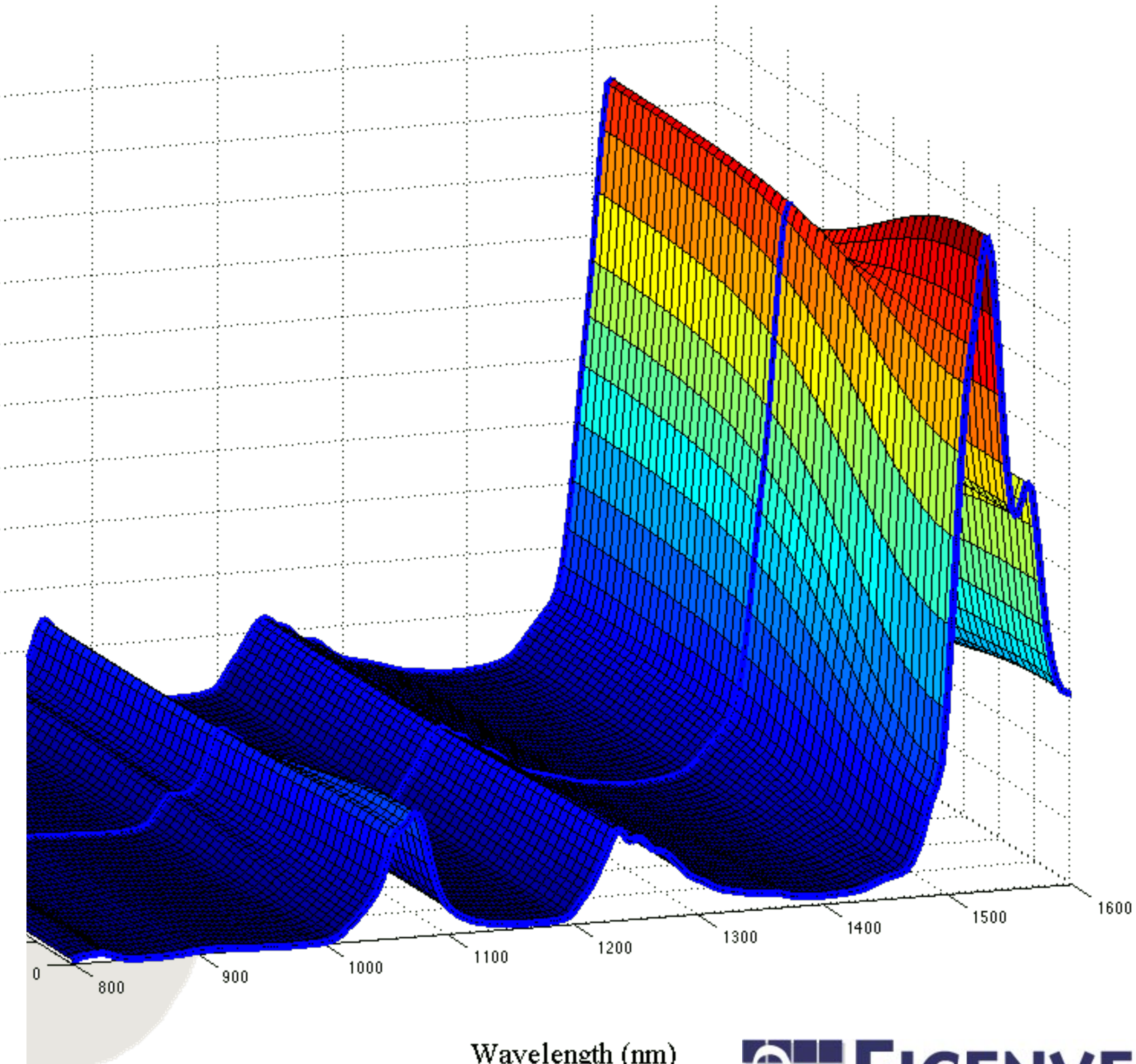


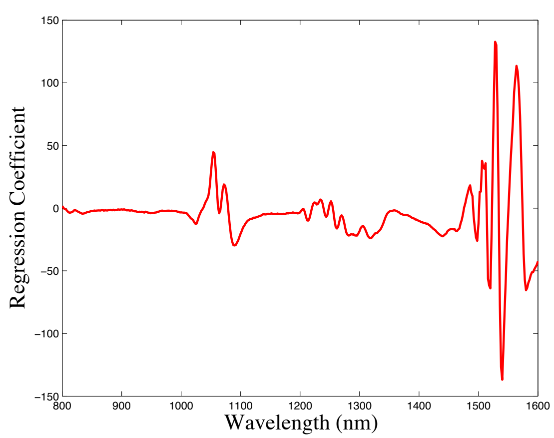
EIGENVECTOR RESEARCH INCORPORATED

Pseudo-gas Scenarios

- Start with orthogonal concentraions
  - 3) All analytes positively correlated
  - 4) One interferents positive, three negative

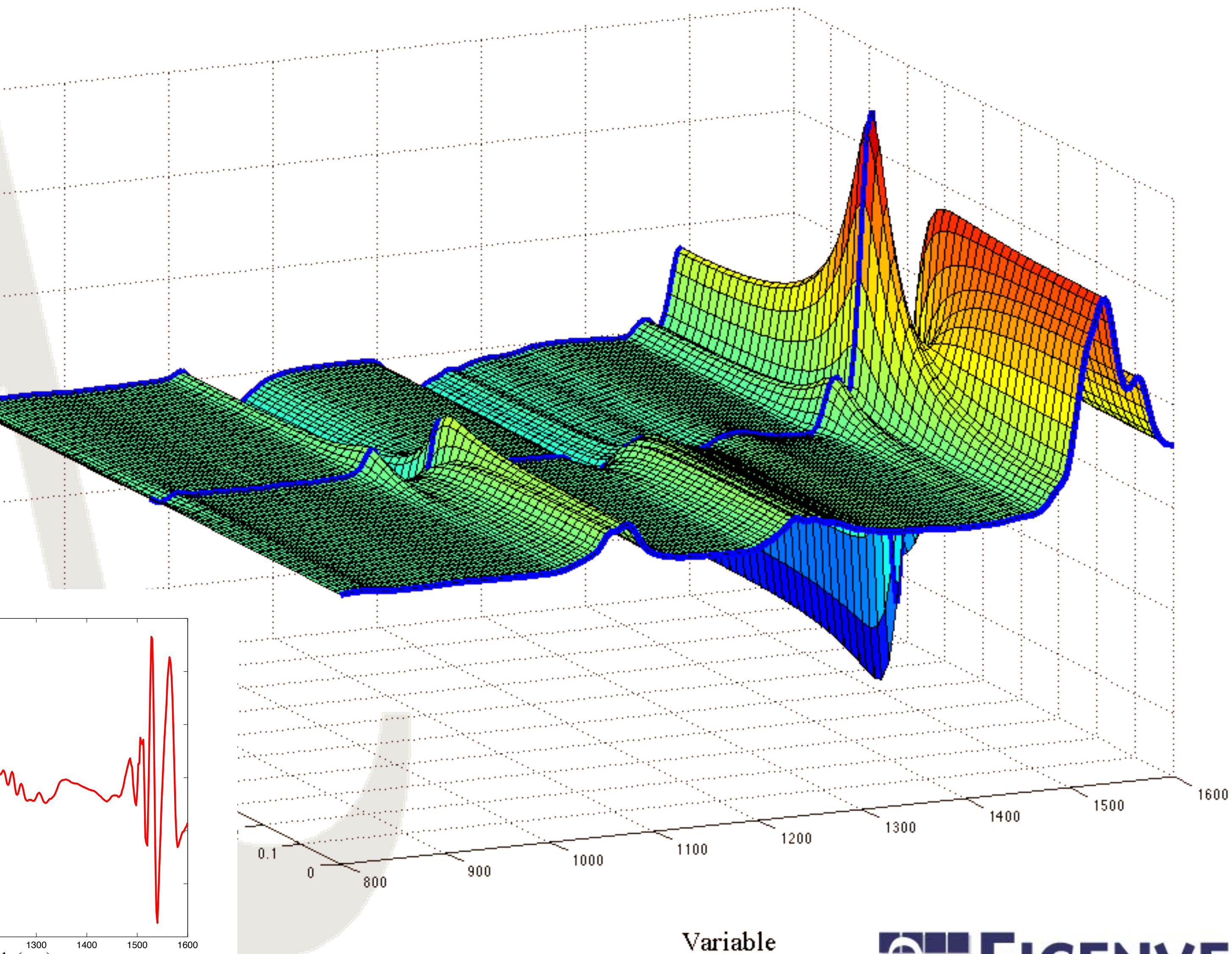
Scenario 3-First O-PLS Loading

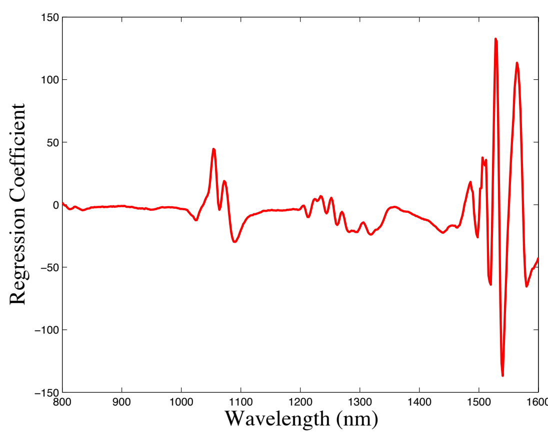




EIGENVECTOR RESEARCH INCORPORATED

Scenario 4-First O-PLS Loading





EIGENVECTOR RESEARCH INCORPORATED

Binary Expression Simulation

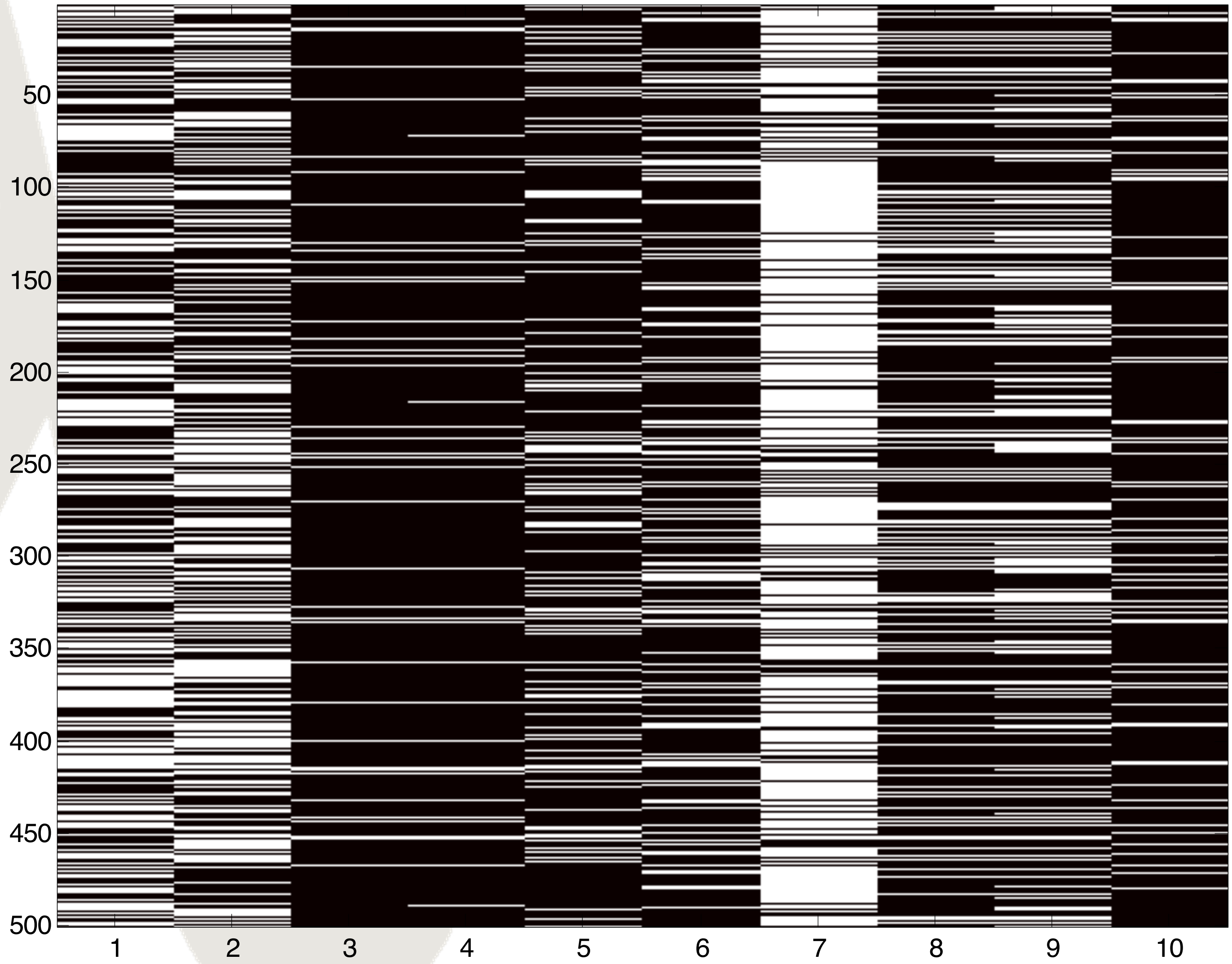
- 10 Expressed Proteins (variables), 500 Subjects
- 1 "primary" effect with loading:

1	2	3	4	5	6	7	8	9	10
☑	☑	☒	☒	0	0	0	0	0	0

  (must have ☑, cannot have ☒, 0 have no effect)

- 7 "background" effects (rank 1 patterns with positive and negative correlations as for primary)
- Only samples with primary loading expression for 1-4 (after mixing all effects) will exhibit property of interest (e.g. disease)

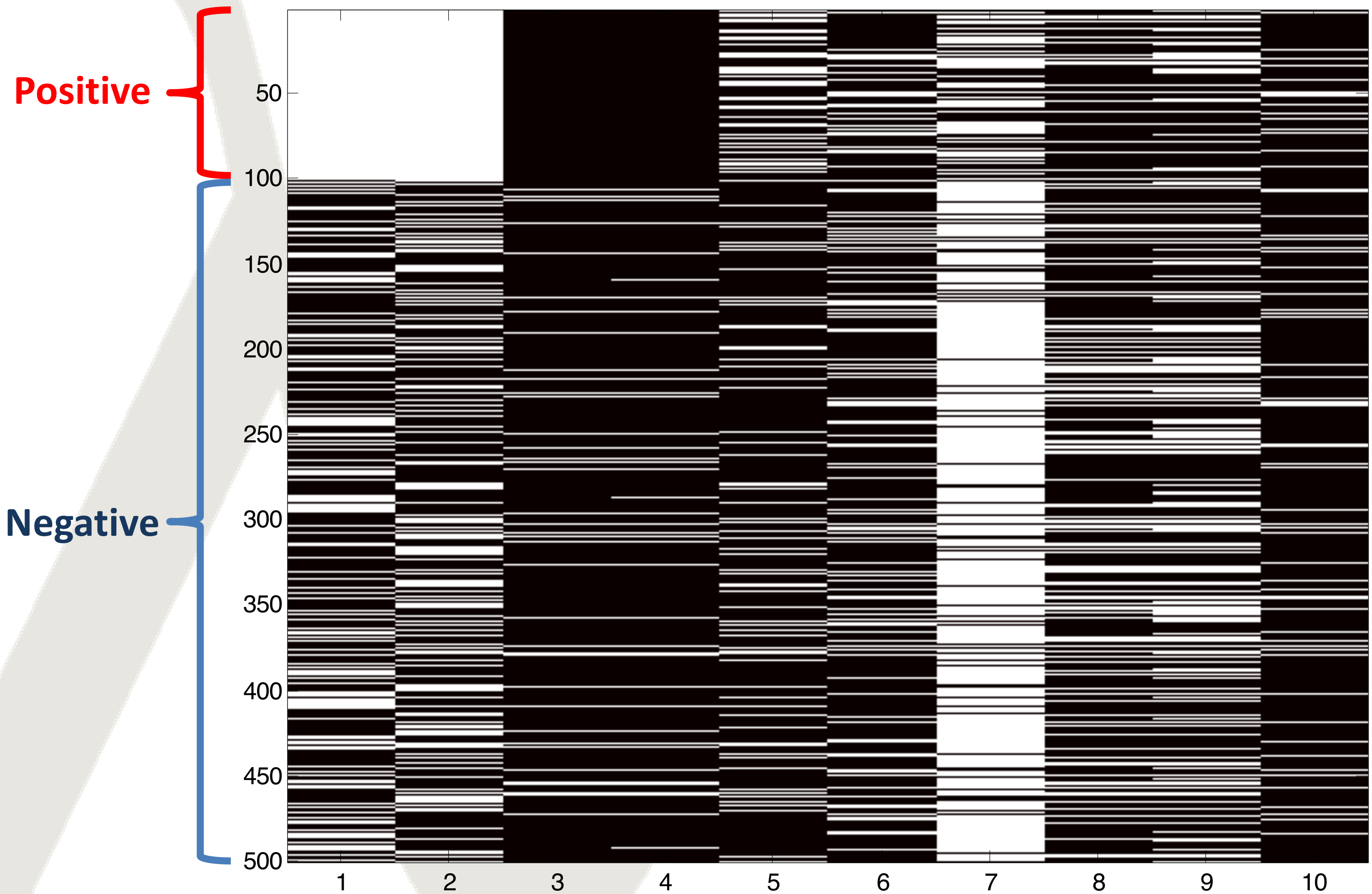
Example Map of Expression



Example Set 1

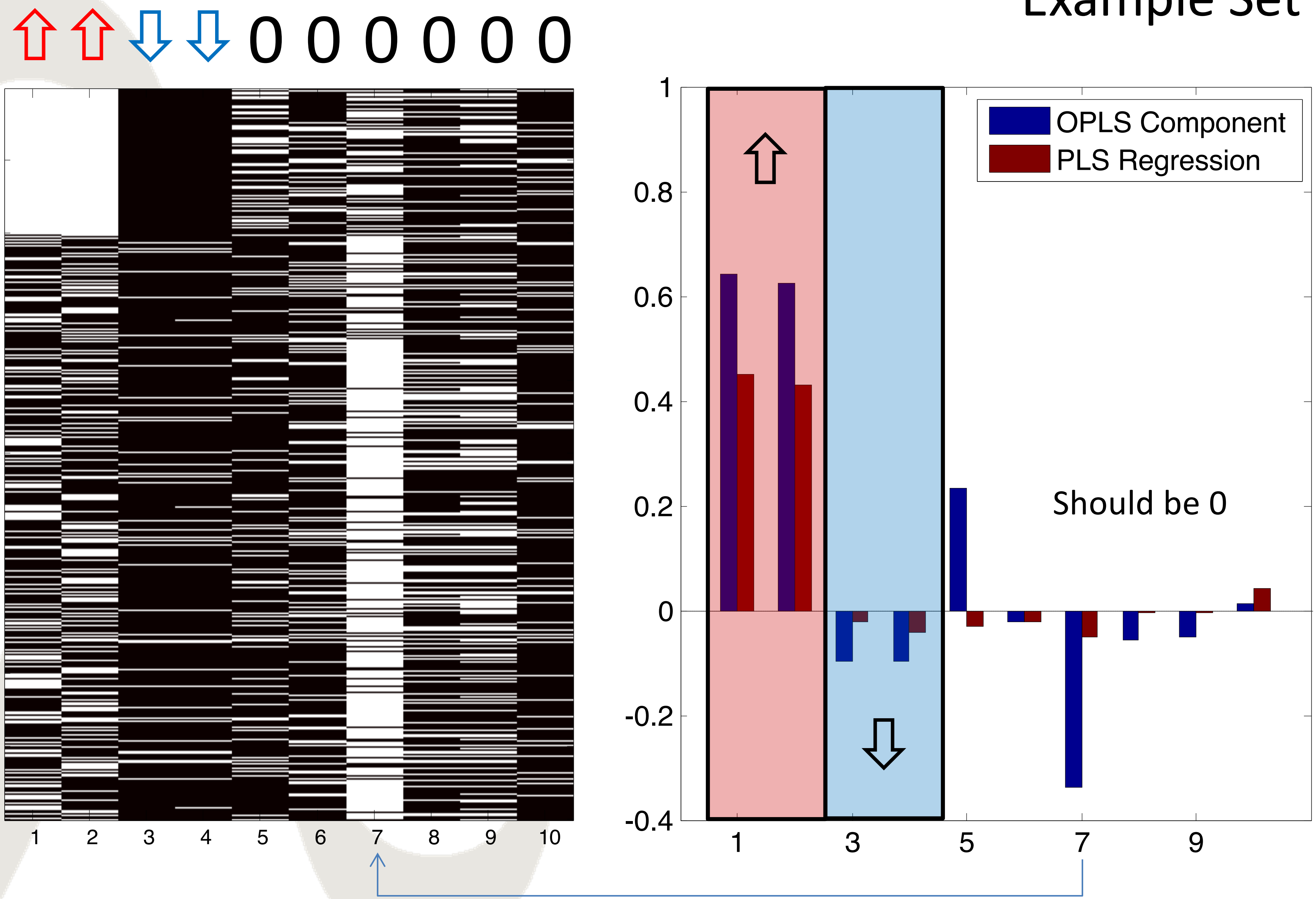
EIGENVECTOR RESEARCH INCORPORATED

...Sorted by Property of Interest

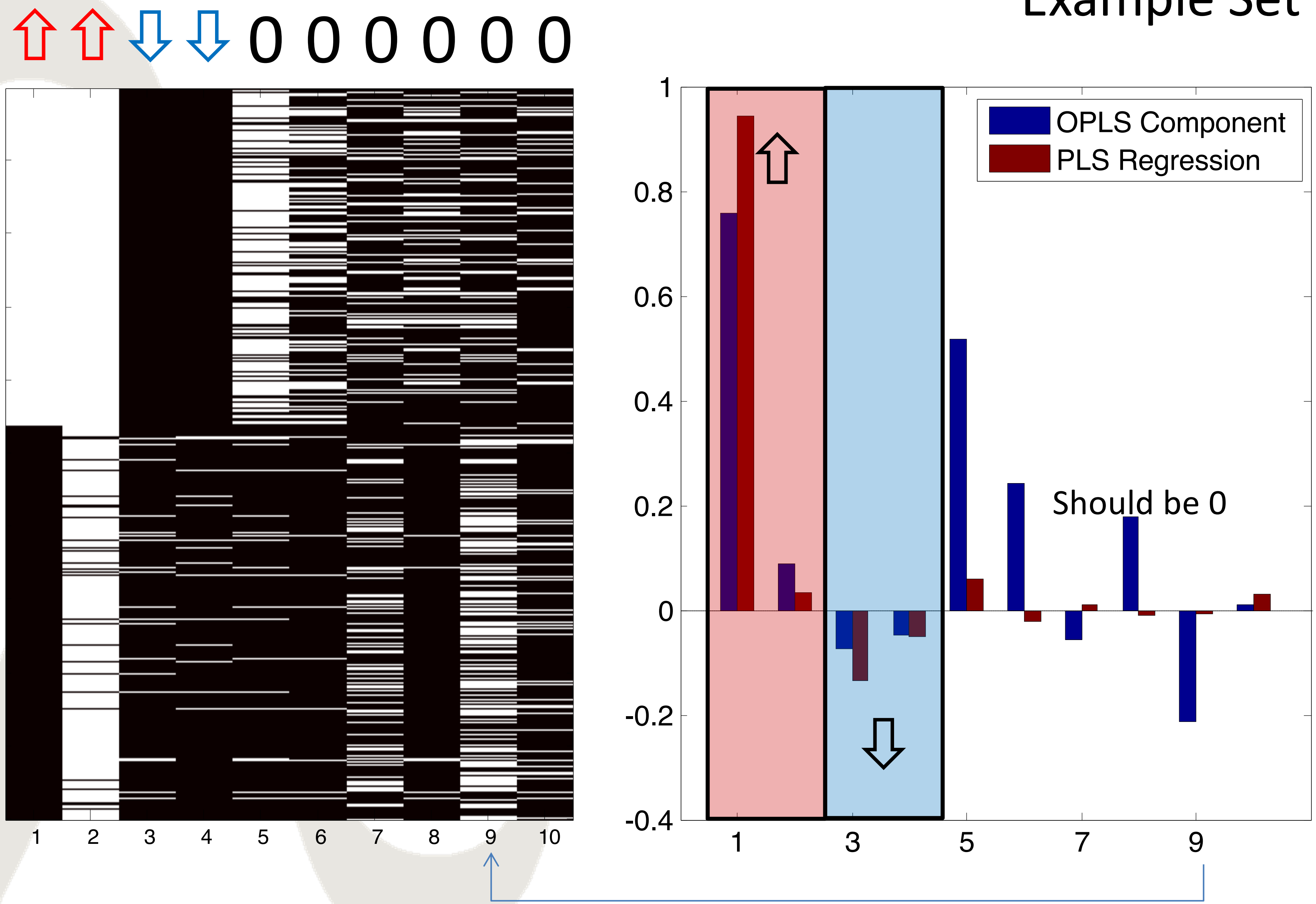


Example Set 1

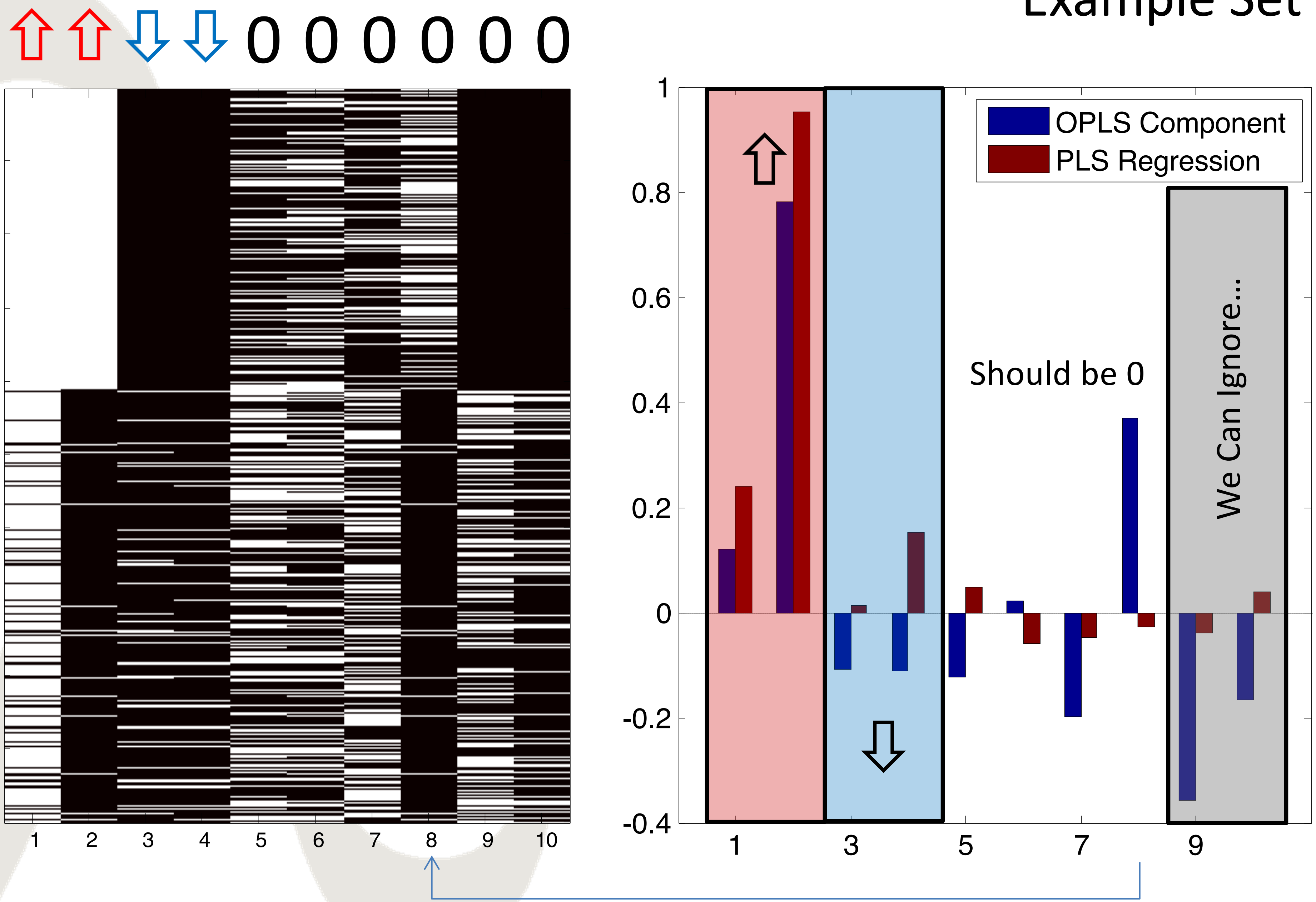
Example Set 1



Example Set 2



Example Set 3



Conclusions

- O-PLS does simplify regression vectors. It is CLOSER to underlying bilinear response…
- … HOWEVER, result generally not the same as a first principles model.
- O-PLS results strong function of correlation in concentrations
- O-PLS recovered component is more sensitive to chance correlation than is regression vector (Problem seen even with 500, 1000, or 2000 samples!)