

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

Analysis - PLS 6 LVs - m5spec, propvals

File Edit Preprocess Analysis Tools Help FigBrowser

- ✓ Cross-Validation
- Orthogonalize Model**
- Show Details
- Report Writer
- Test Model Robustness
- Permutation Test
- Correlation Map
- Estimate Factor SNR
- View Cache
- Toolbar

View: SSQ Table

Number LVs: 6 Auto Select

Latent Variable	Percent Variance X-Block		Percent Variance Y-Block	
	LV	Cum	LV	Cum
1	99.08	99.08	39.05	39.05
2	0.76	99.84	19.26	58.31
3	0.06	99.90	23.49	81.79
4	0.03	99.93	14.25	96.04
5	0.03	99.96	2.24	98.28
6	0.01	99.98	1.00	99.28 ← Suggested
7	0.01	99.98	0.31	99.59
8	0.01	99.99	0.09	99.68
9	0.00	99.99	0.16	99.83
10	0.00	100.00	0.02	99.85
11	0.00	100.00	0.09	99.94
12	0.00	100.00	0.02	99.96
13	0.00	100.00	0.01	99.97
14	0.00	100.00	0.00	99.97
15	0.00	100.00	0.00	99.98
16	0.00	100.00	0.01	99.98



Analysis - PLS 6 LVs - m5spec, propvals

File Edit Preprocess Analysis Tools Help FigBrowser

- ✓ Cross-Validation
- Orthogonalize Model** ✓
- Show Details
- Report Writer
- Test Model Robustness
- Permutation Test
- Correlation Map
- Estimate Factor SNR
- View Cache
- Toolbar

View: SSQ Table iPLS Variable Selection

Number LVs: 6 Auto Select

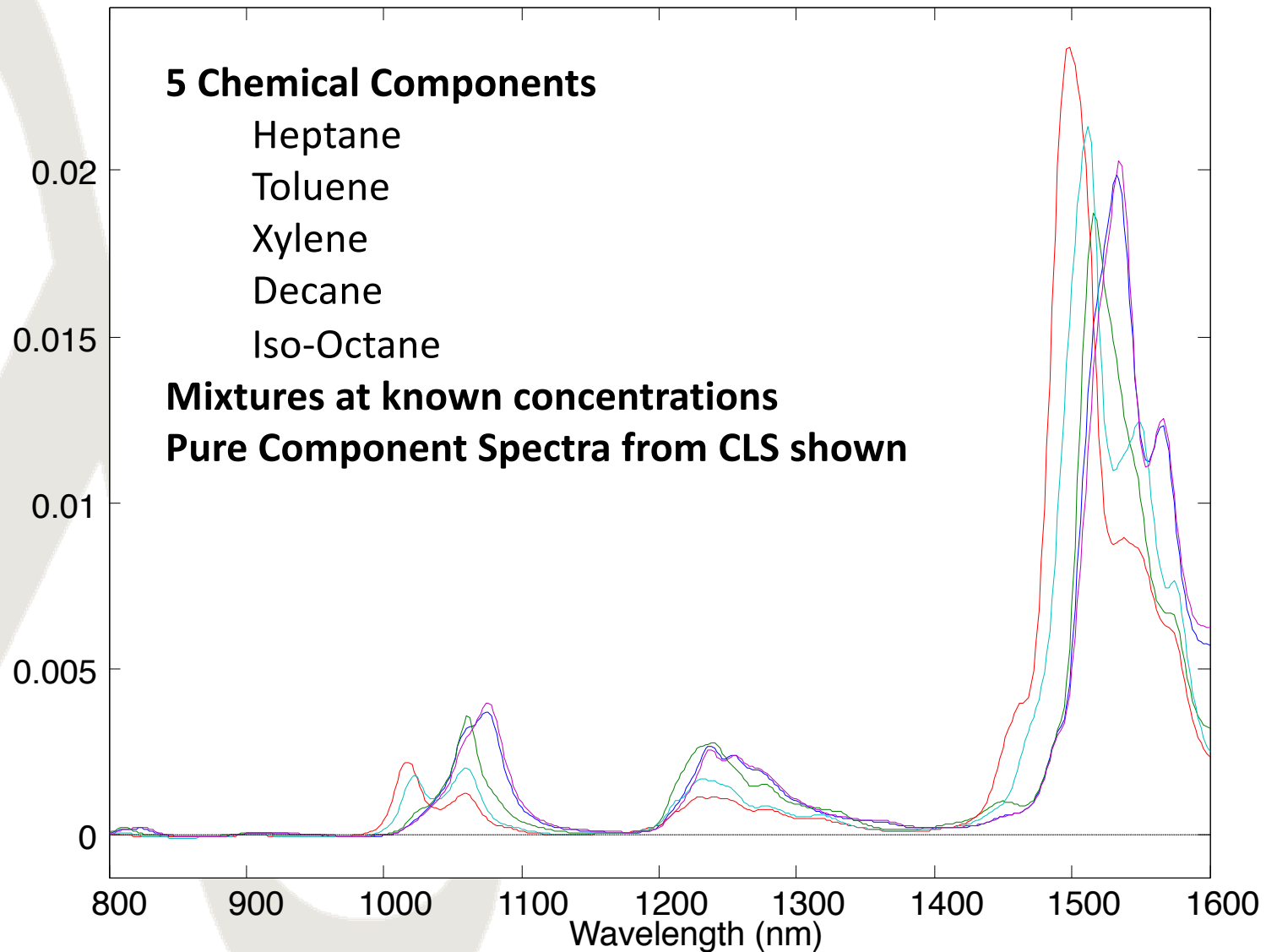
Latent Variable	Percent Variance Captured by Model X-Block		Percent Variance Captured by Model Y-Block	
	LV	Cum	LV	Cum
1	96.86	96.86	99.28	99.28
2	2.73	99.60	0.00	99.28
3	0.20	99.80	0.00	99.28
4	0.10	99.90	0.00	99.28
5	0.06	99.95	0.00	99.28
6	0.02	99.98	0.00	99.28 ← Suggested
7	0.01	99.98	0.31	99.59
8	0.01	99.99	0.09	99.68
9	0.00	99.99	0.16	99.83
10	0.00	100.00	0.02	99.85
11	0.00	100.00	0.09	99.94
12	0.00	100.00	0.02	99.96
13	0.00	100.00	0.01	99.97
14	0.00	100.00	0.00	99.97
15	0.00	100.00	0.00	99.98
16	0.00	100.00	0.01	99.98

Questions:

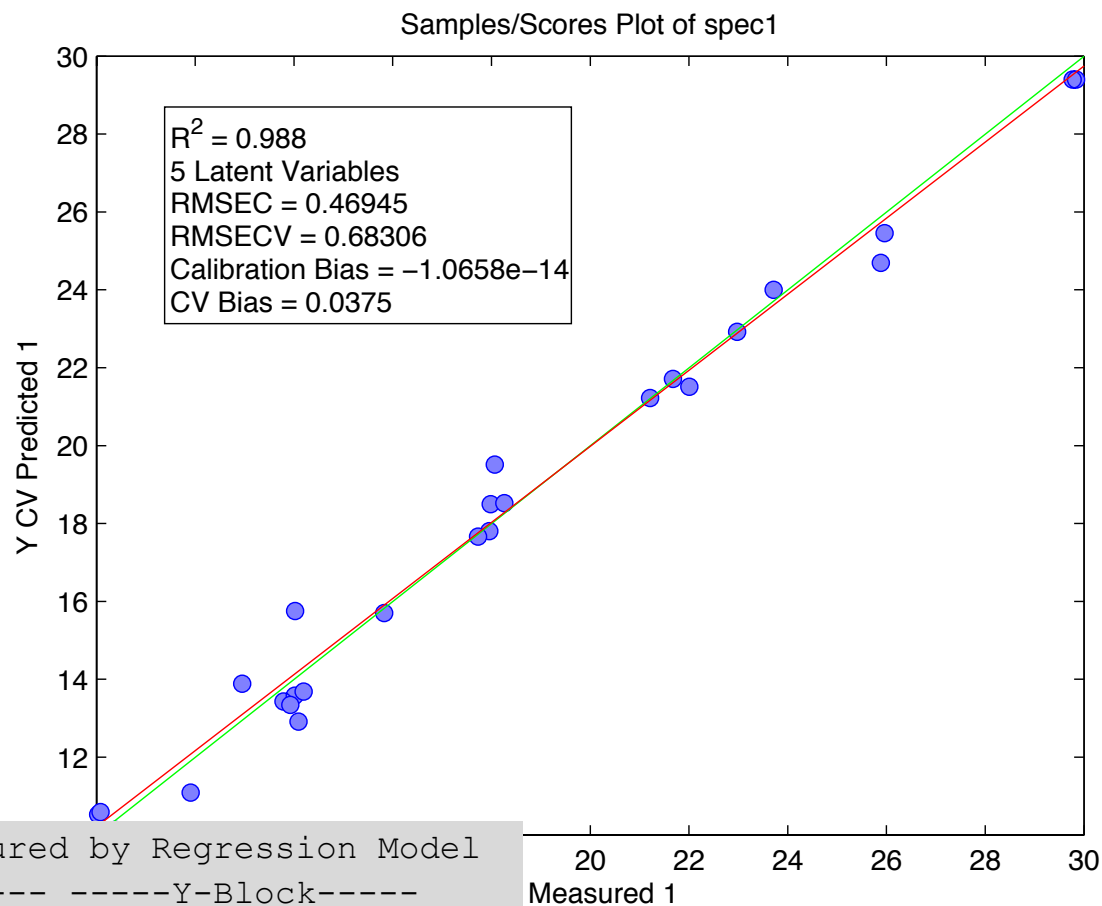
- What do we need to be aware of when interpreting O-PLS recovered components?
- What kinds of sensitivities does O-PLS have to noise, rotational ambiguity, and correlated concentrations?

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

NIR of Pseudo-gasoline Samples



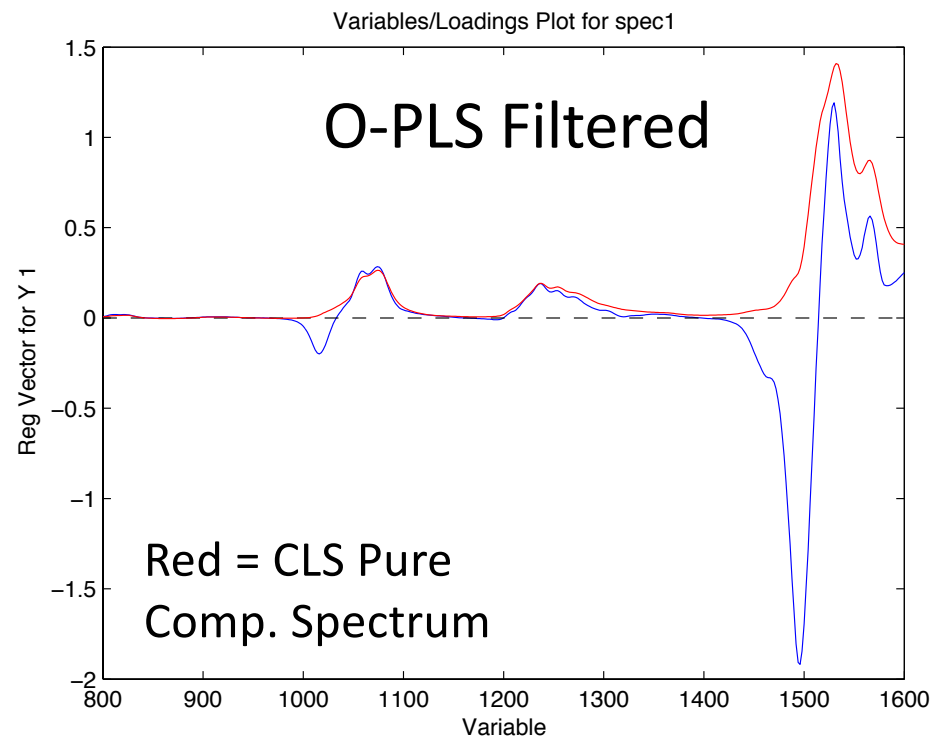
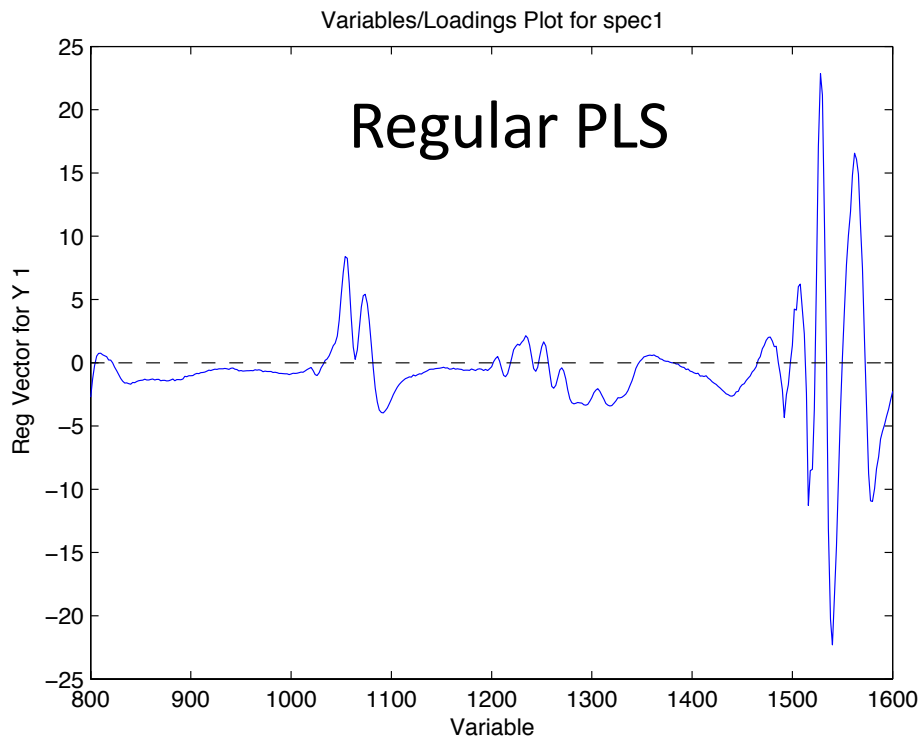
PLS Model of Heptane



Percent Variance Captured by Regression Model

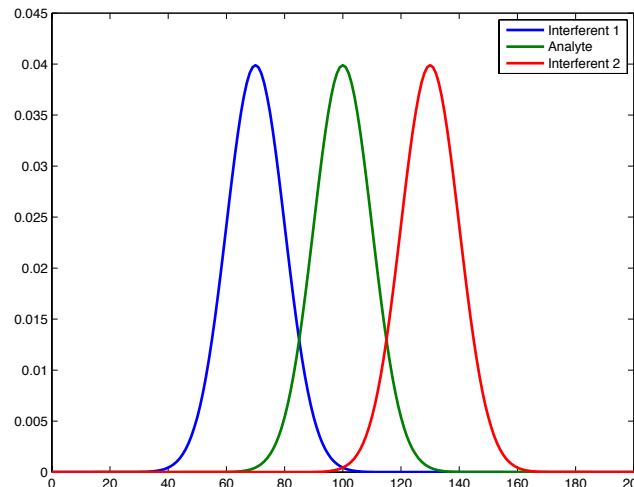
Comp	-----X-Block-----		-----Y-Block-----	
	This	Total	This	Total
1	91.17	91.17	8.36	8.36
2	7.40	98.57	7.19	15.55
3	0.93	99.50	32.81	48.36
4	0.46	99.96	26.18	74.54
5	0.02	99.98	24.90	99.44

Regular PLS and O-PLS Filtered Regression Vectors



Simple System Example

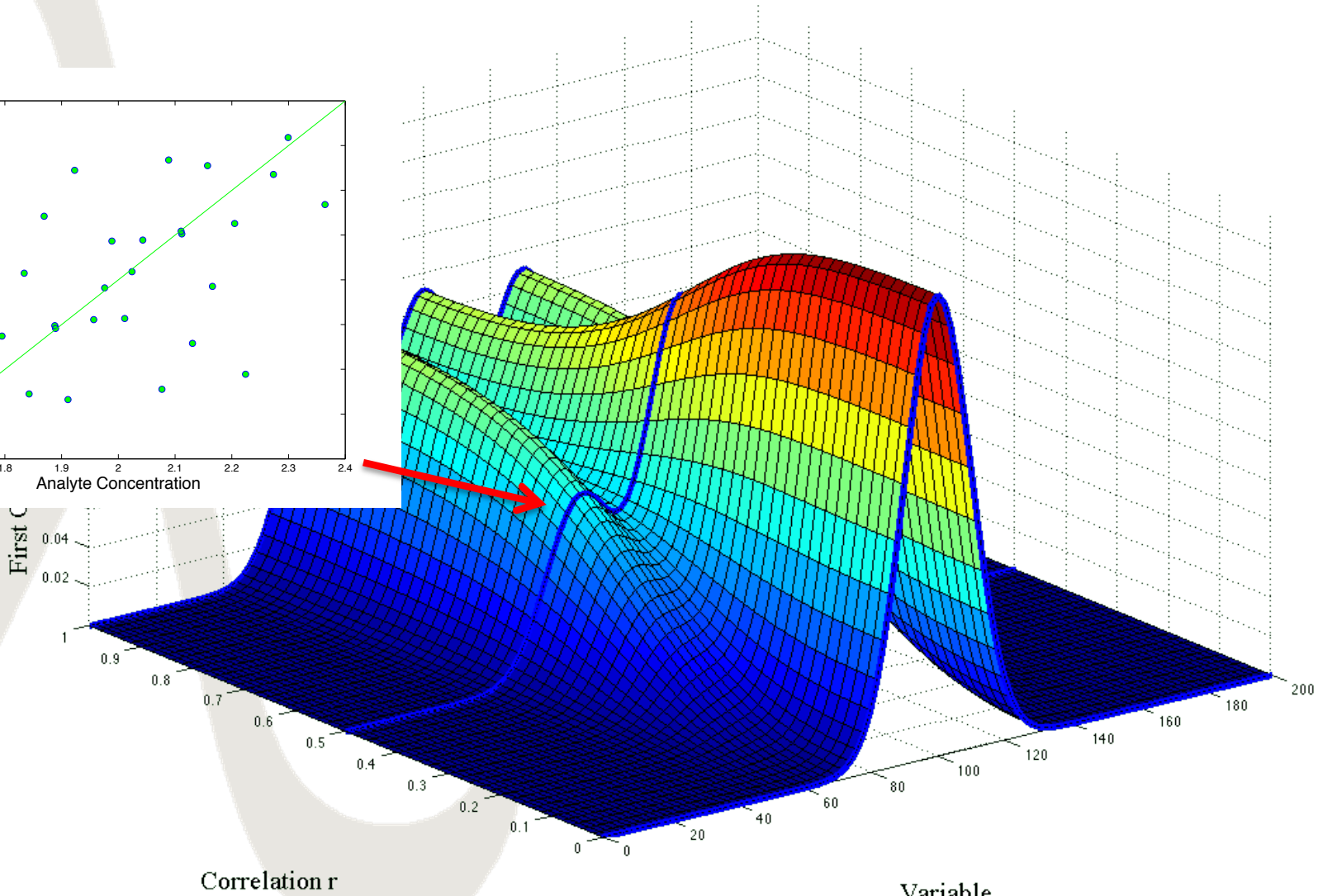
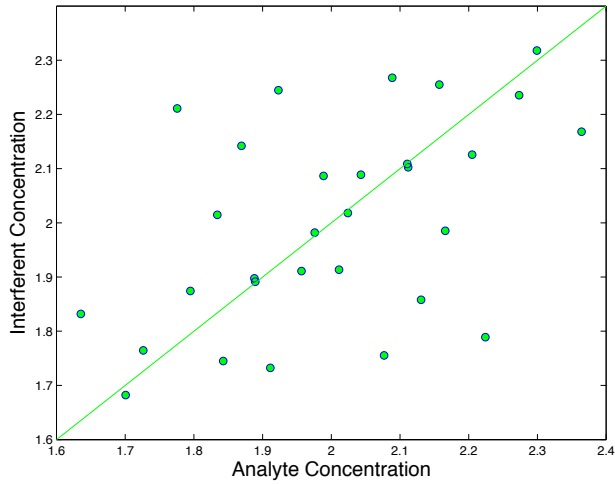
- Synthetic example of three constituents
- Evenly spaced Gaussian peaks, analyte in middle
- Vary correlation between analyte and interferences



Gaussian Peaks Scenarios

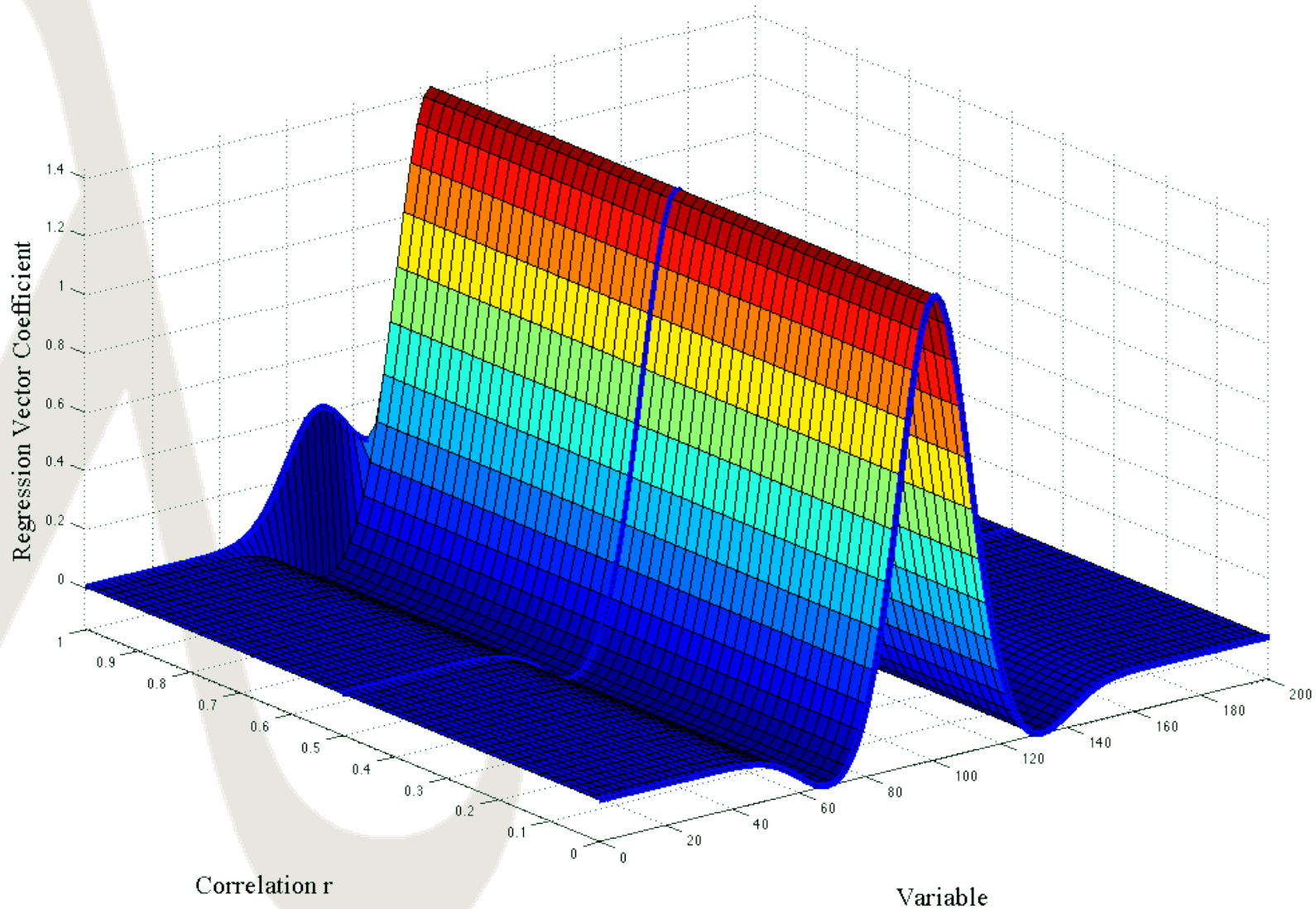
- Start with orthogonal concentrations
 - 1) Go from orthogonal to positively correlated concentrations
 - 2) One interferent positively correlated, one negatively correlations

Scenario 1-First O-PLS Loading

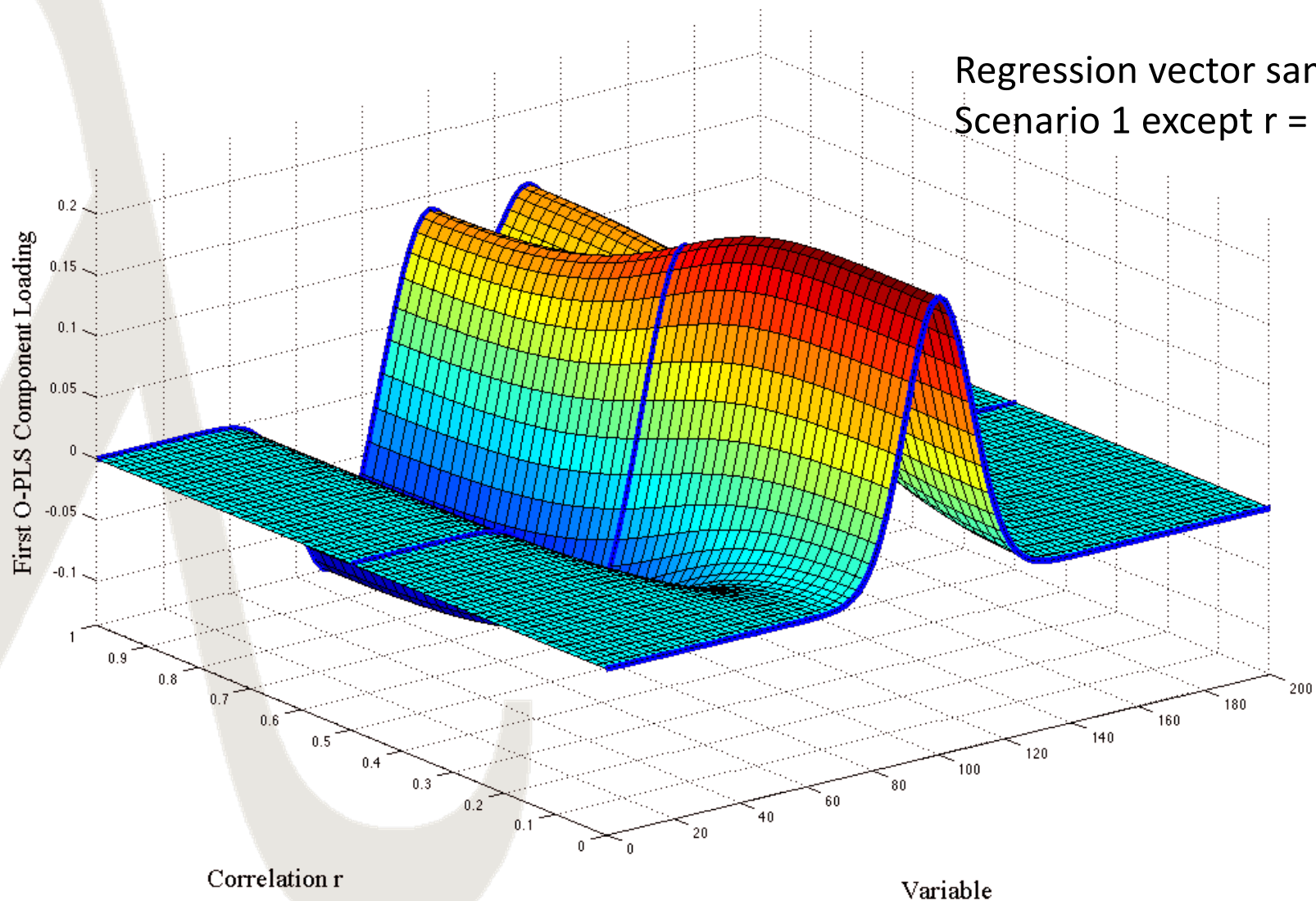


Variable

Scenario 1-Regression Vector



Scenario 2-First O-PLS Loading

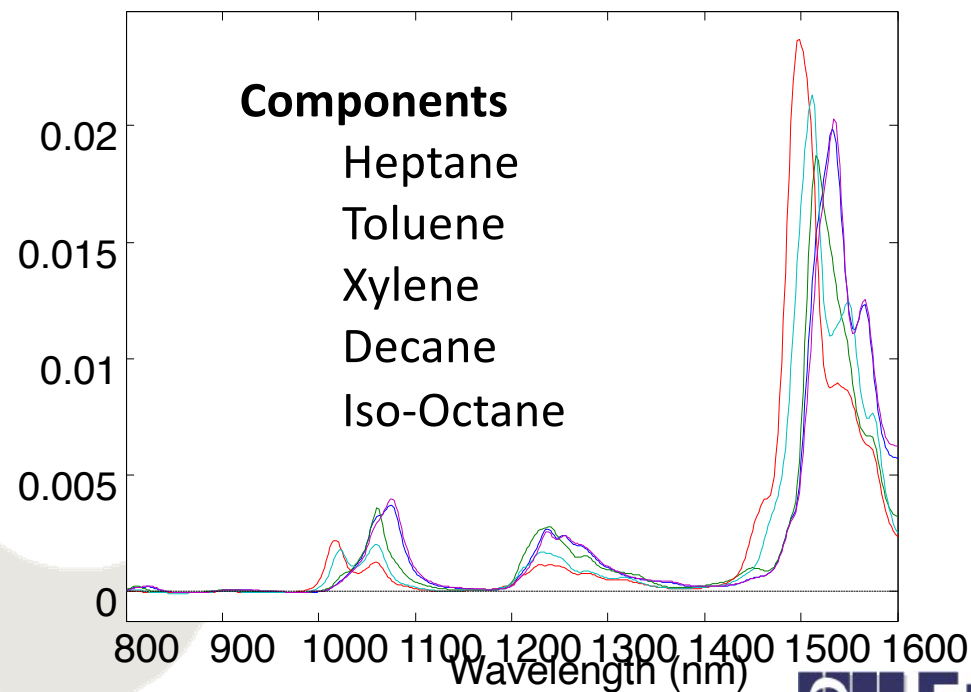


Regression vector same as Scenario 1 except $r = 1$ case

Variable

Pseudo-gasoline Example

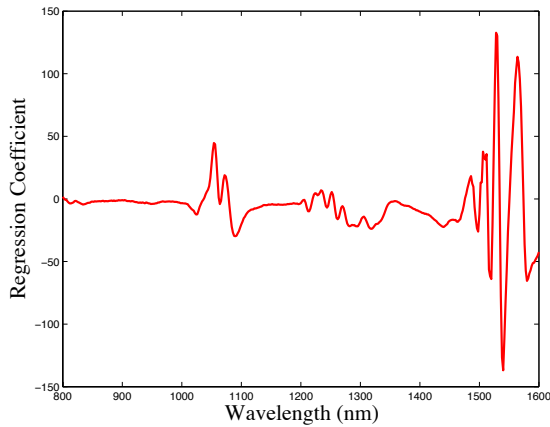
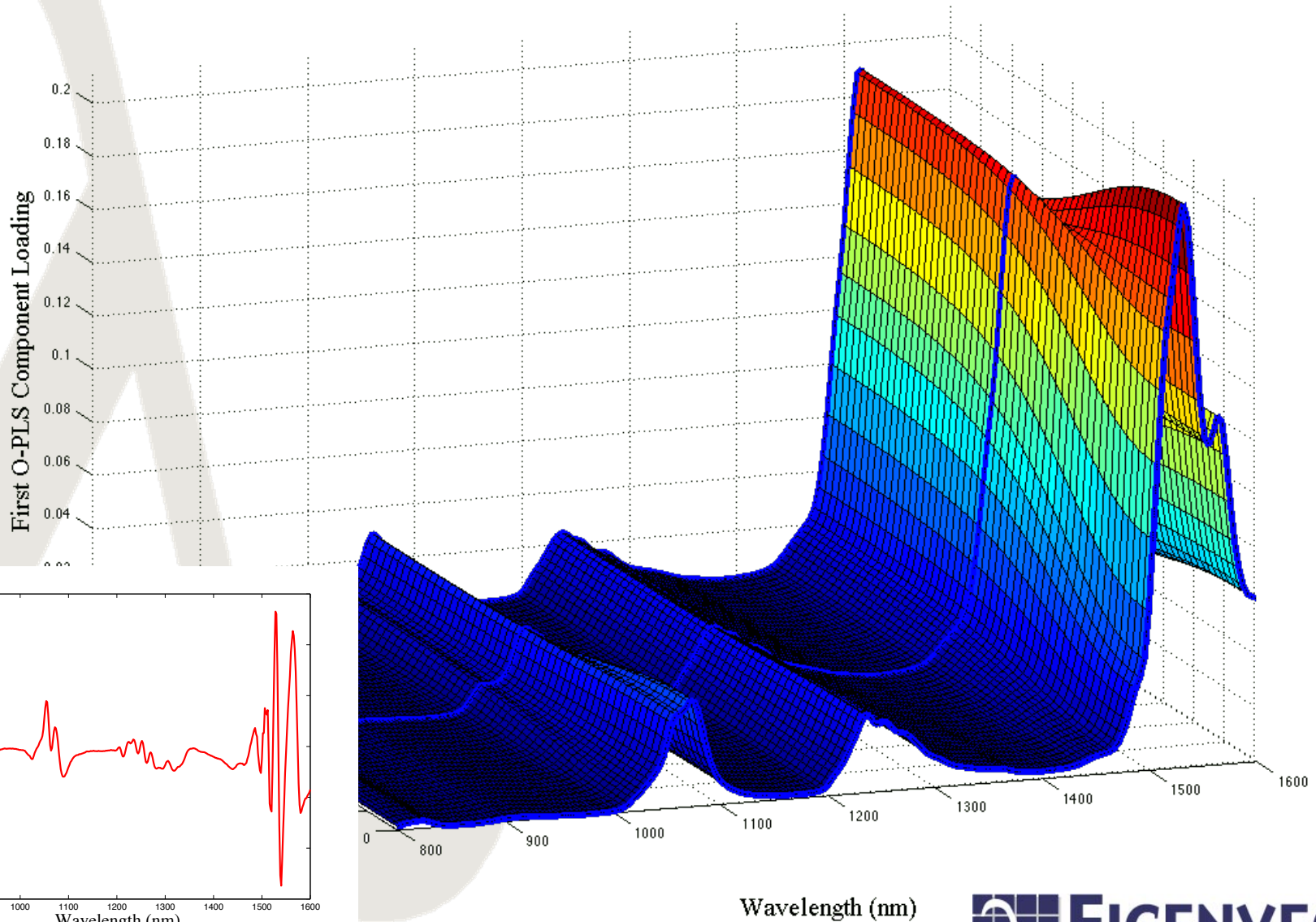
- 5 component mixture measured by NIR
- Solve for pure components via CLS
- Use pure spectra to create synthetic scenarios



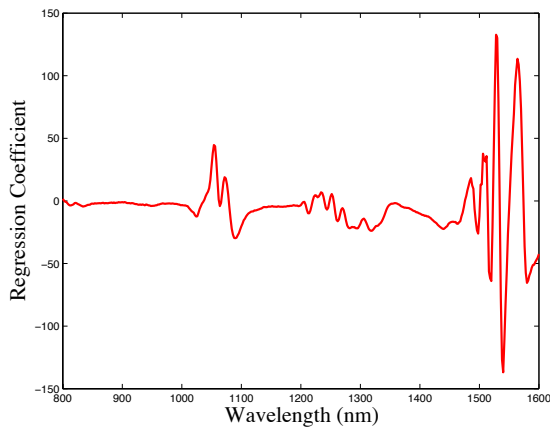
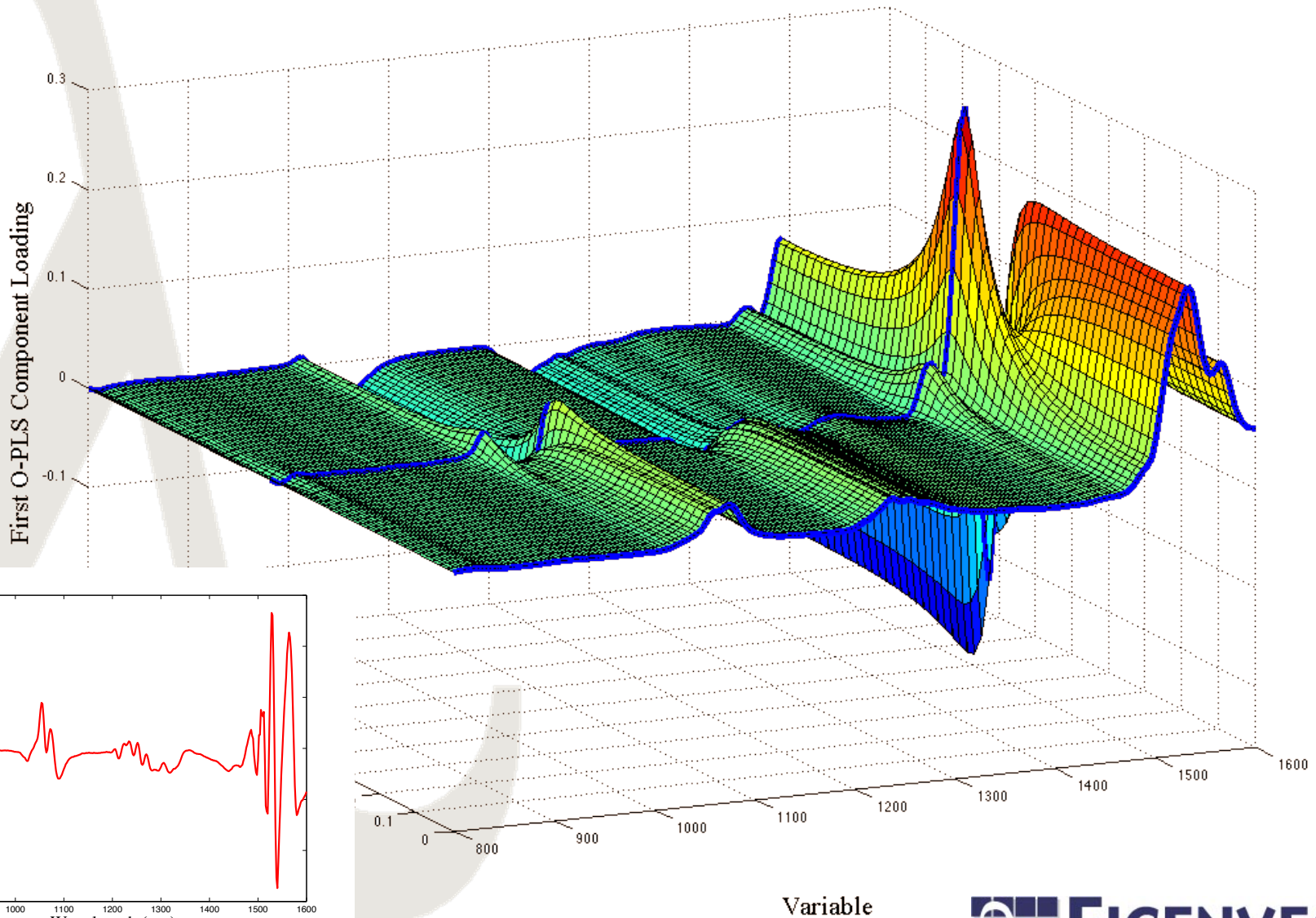
Pseudo-gas Scenarios

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

Scenario 3-First O-PLS Loading



Scenario 4-First O-PLS Loading



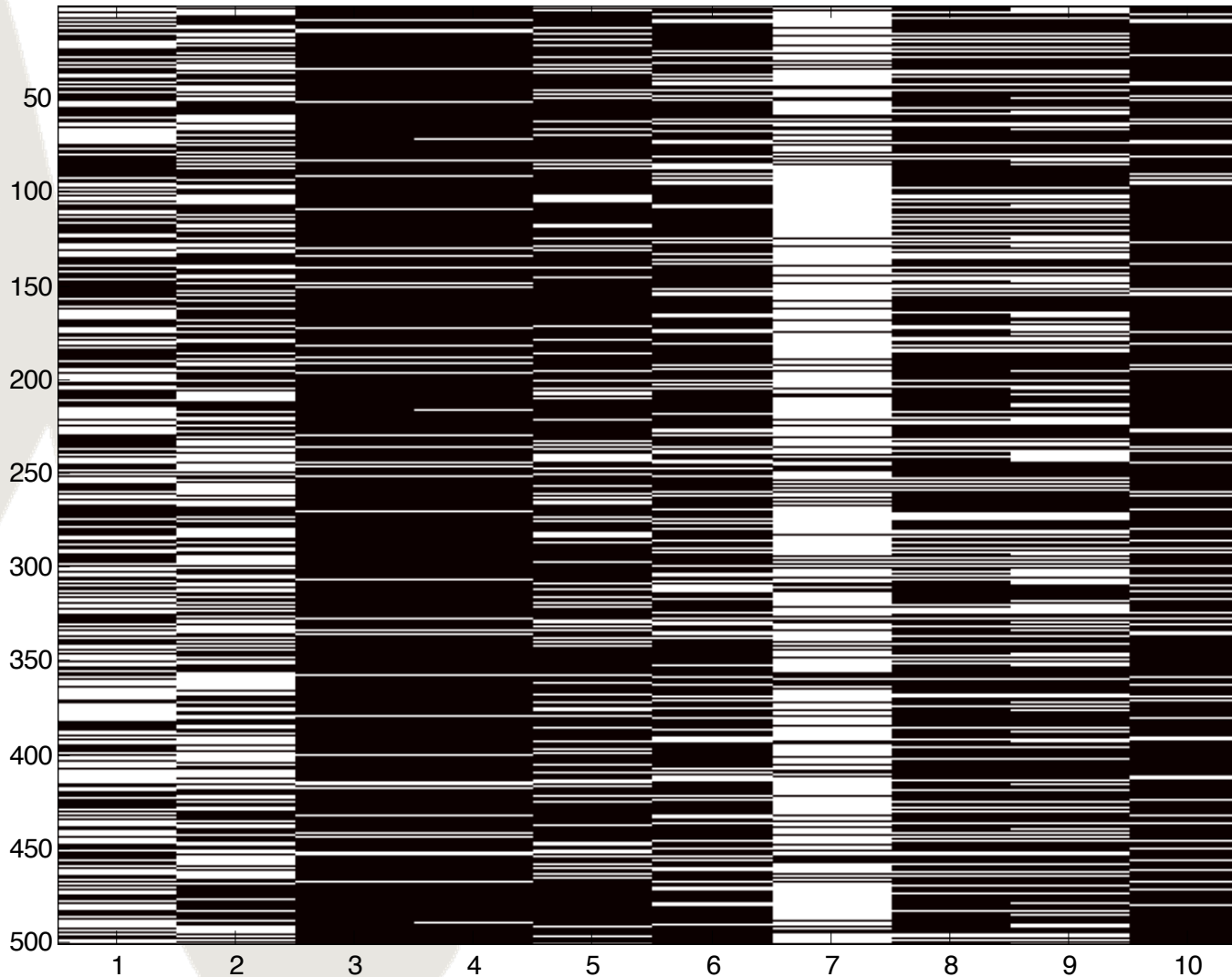
Binary Expression Simulation

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

1	2	3	4	5	6	7	8	9	10
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0	0	0	0	0	0

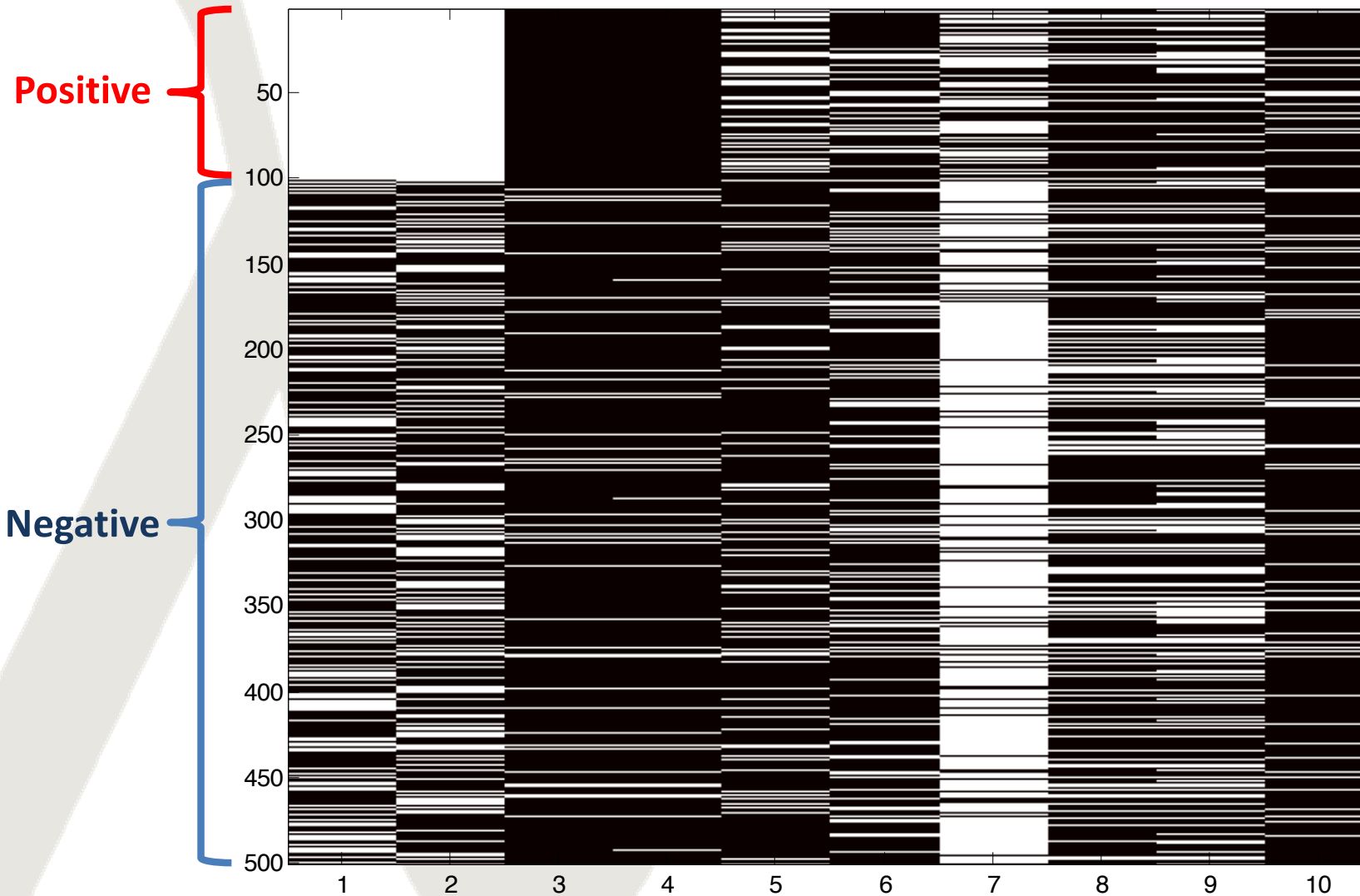
(must have , cannot have , 0 have no effect)
- 7 "background" effects (rank 1 patterns with positive and negative correlations as for primary)
- Only samples with primary loading expression for 1-4 (after mixing all effects) will exhibit property of interest (e.g. disease)

Example Map of Expression



Example Set 1

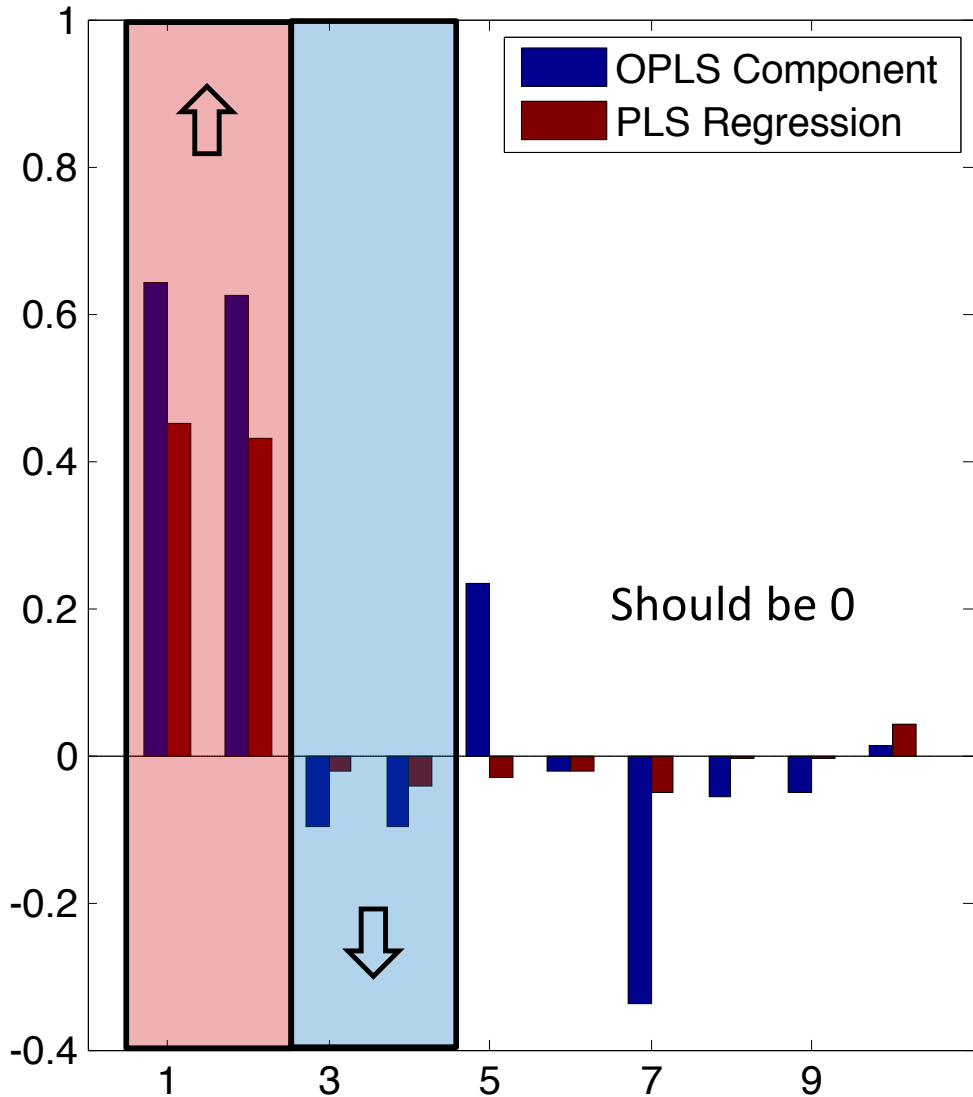
...Sorted by Property of Interest



Example Set 1

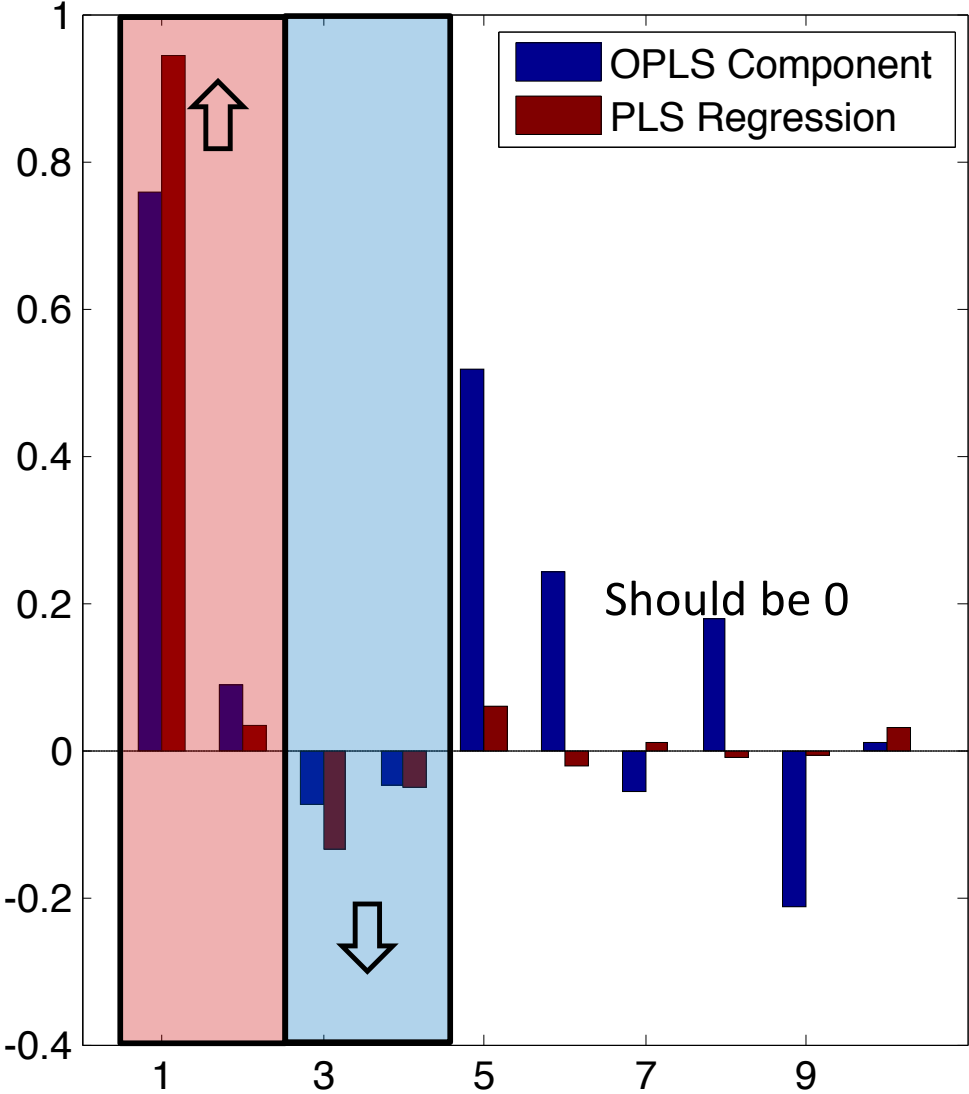
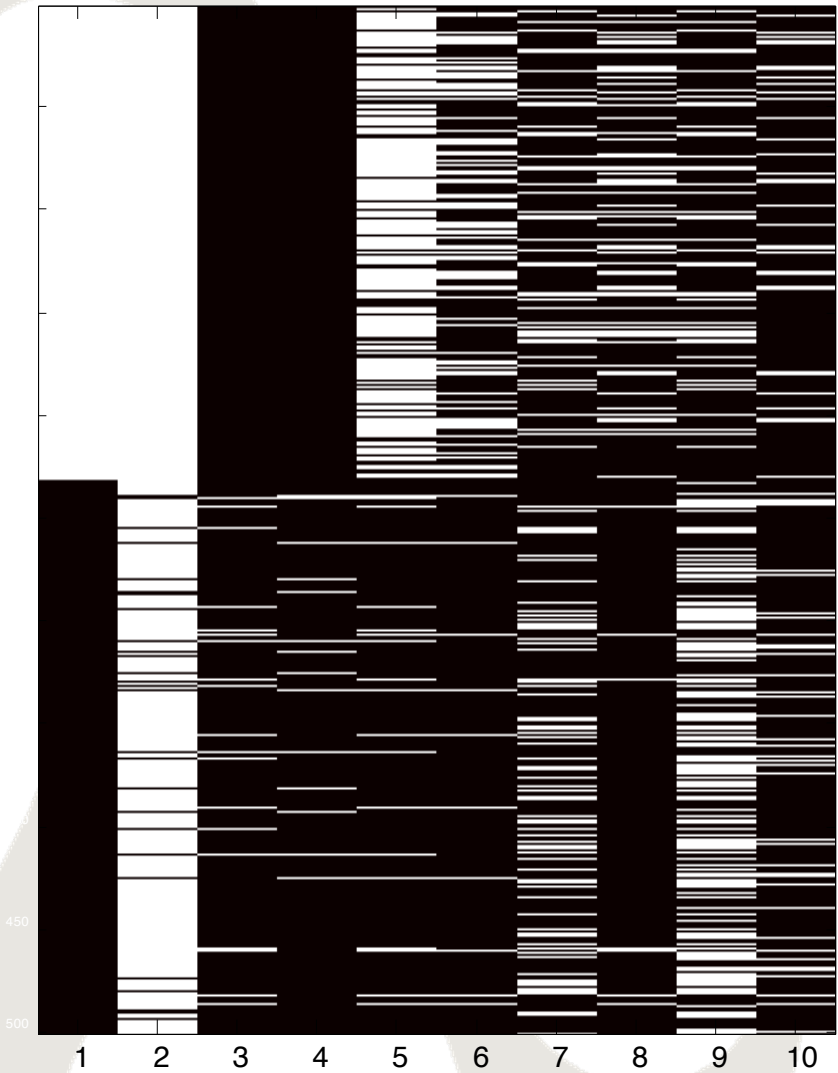
Example Set 1

↑ ↑ ↓ ↓ 0 0 0 0 0 0



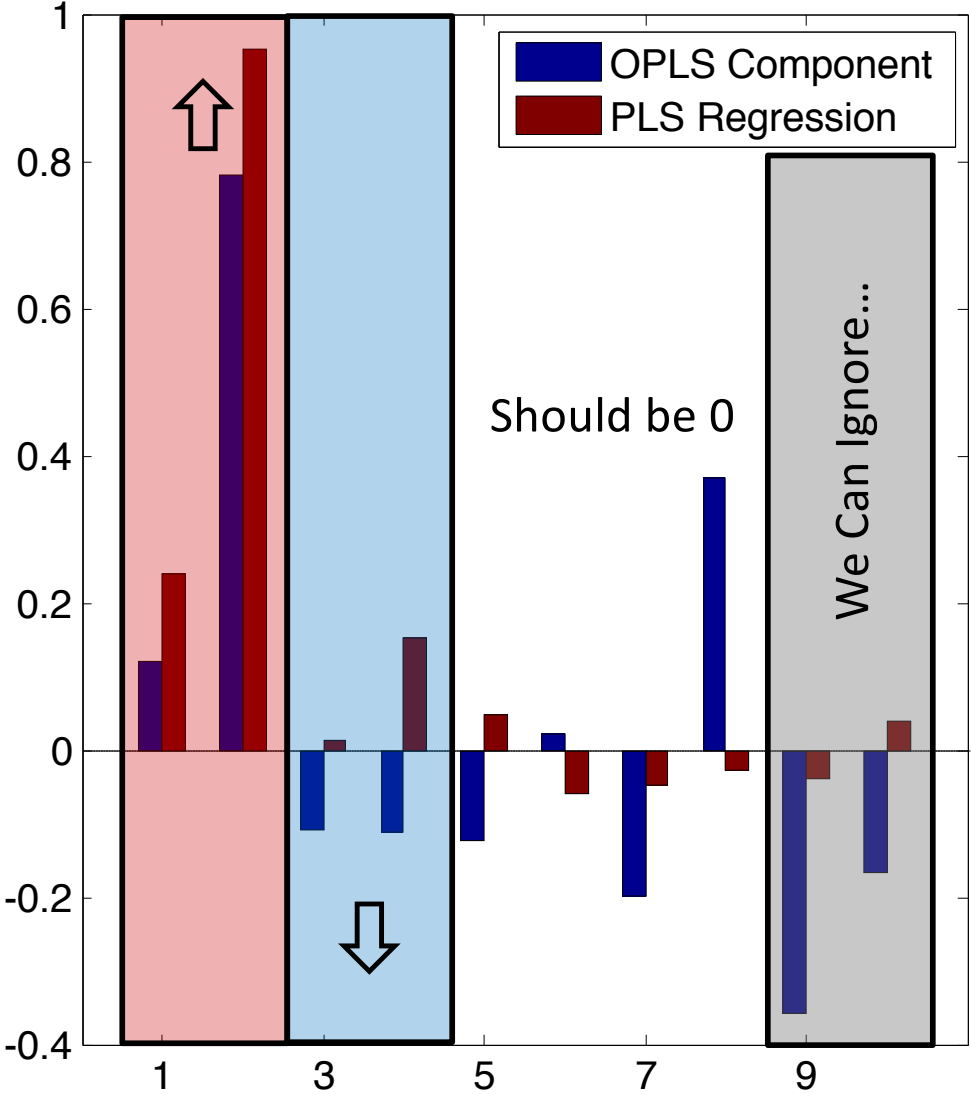
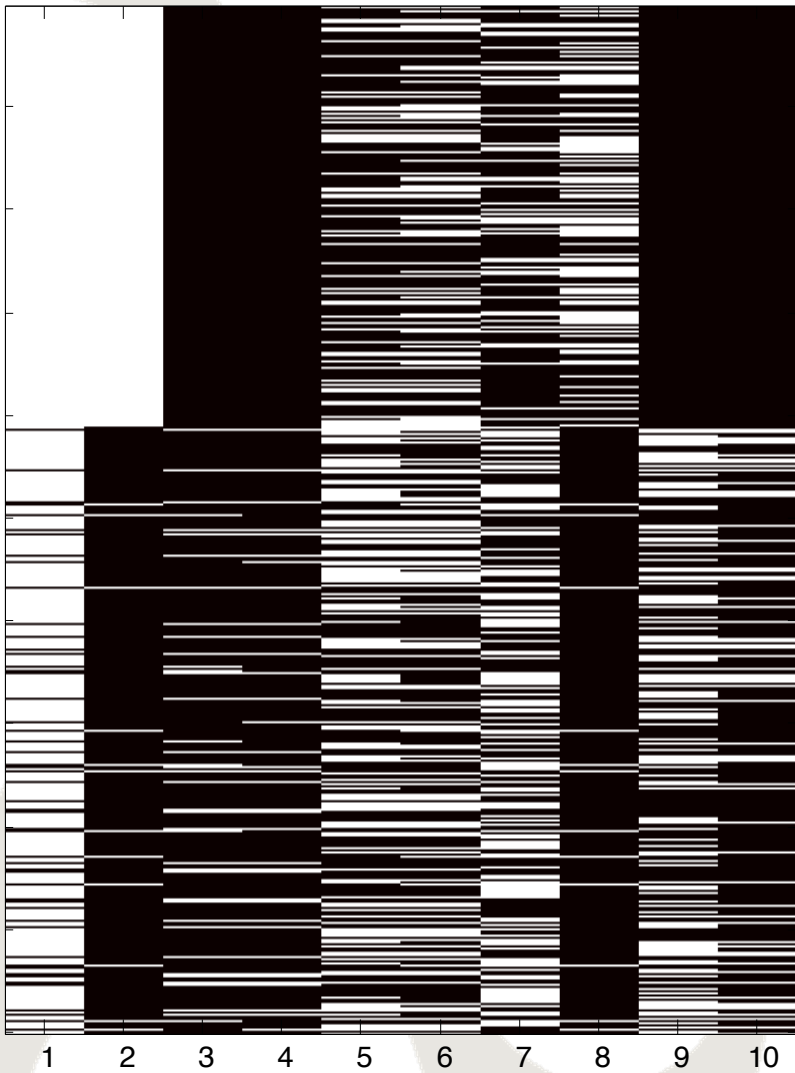
Example Set 2

↑ ↑ ↓ ↓ 0 0 0 0 0 0



Example Set 3

↑ ↑ ↓ ↓ 0 0 0 0 0 0



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!)