On the Interpretability of O-PLS Filtered Models

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

EIGENVECTOR RESEARCH INCORPORATED

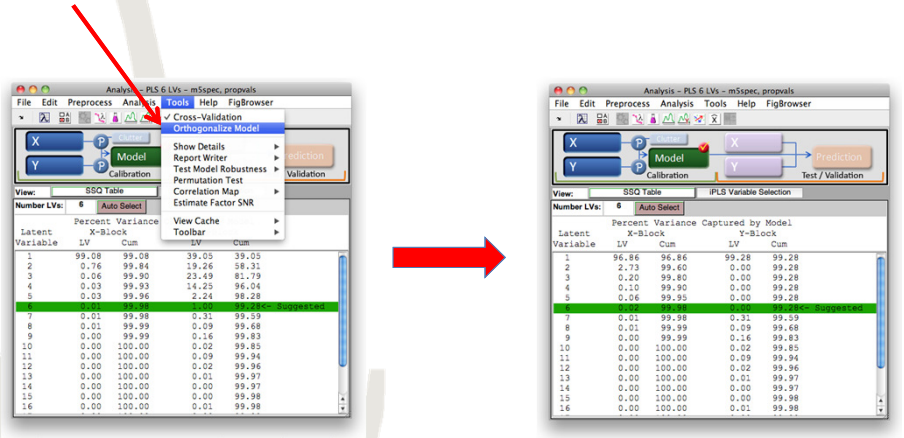
O-PLS

- 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.

EIGENVECTOR RESEARCH INCORPORATED

Orthogonalize Model



EIGENVECTOR RESEARCH INCORPORATED

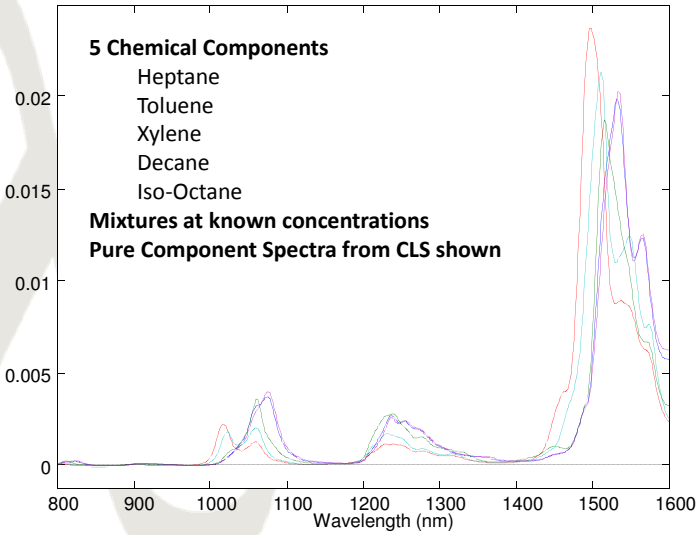
Questions:

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

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

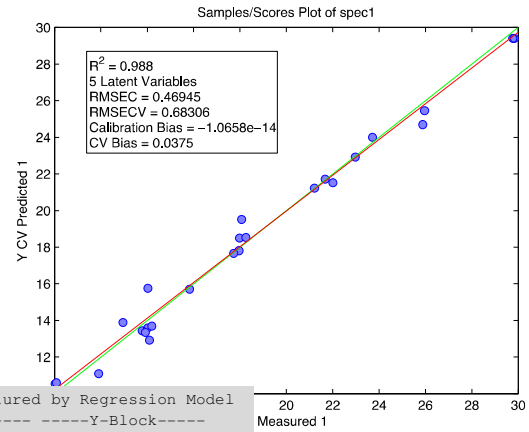
EIGENVECTOR RESEARCH INCORPORATED

NIR of Pseudo-gasoline Samples



EIGENVECTOR RESEARCH INCORPORATED

PLS Model of Heptane

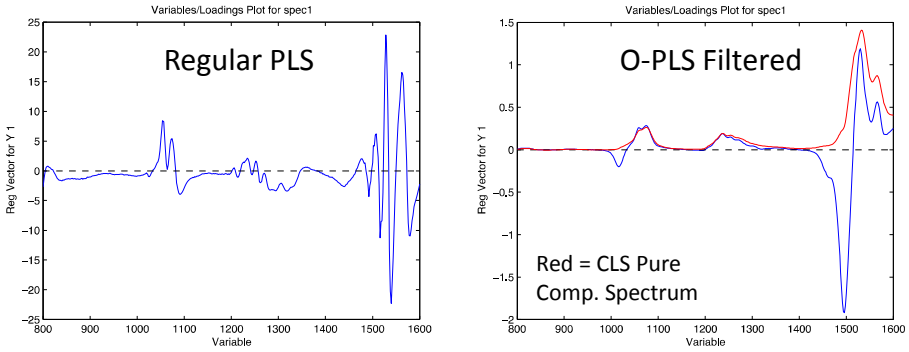


Percent Variance Captured by Regression Model

	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

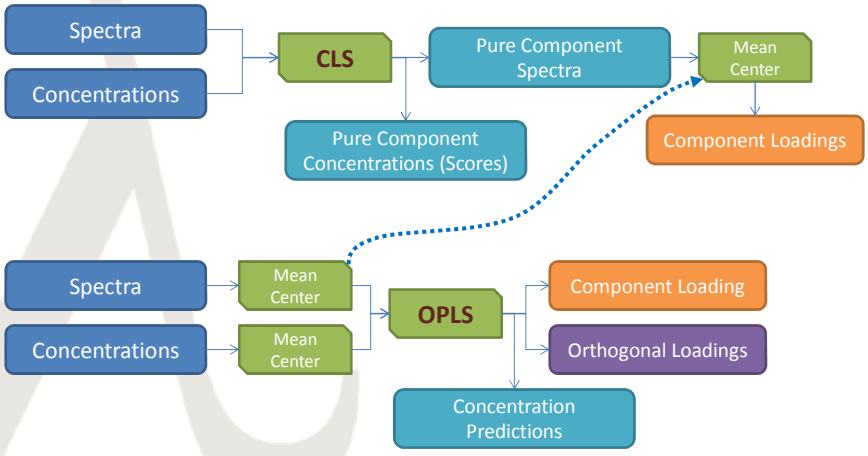
EIGENVECTOR RESEARCH INCORPORATED

Regular PLS and O-PLS Filtered Regression Vectors



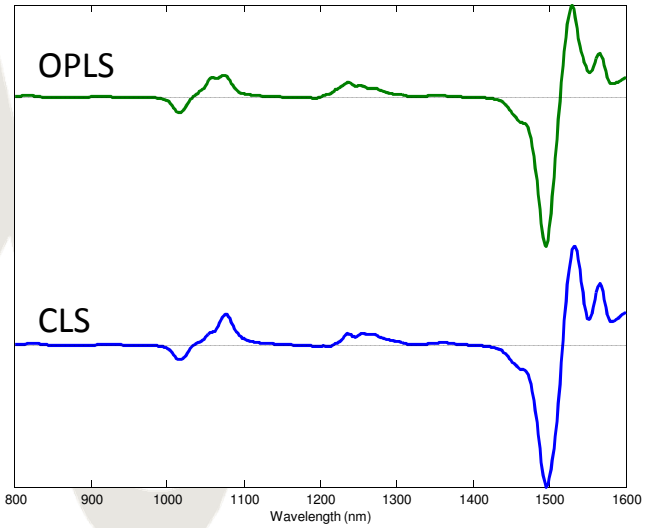
EIGENVECTOR RESEARCH INCORPORATED

OPLS vs. CLS



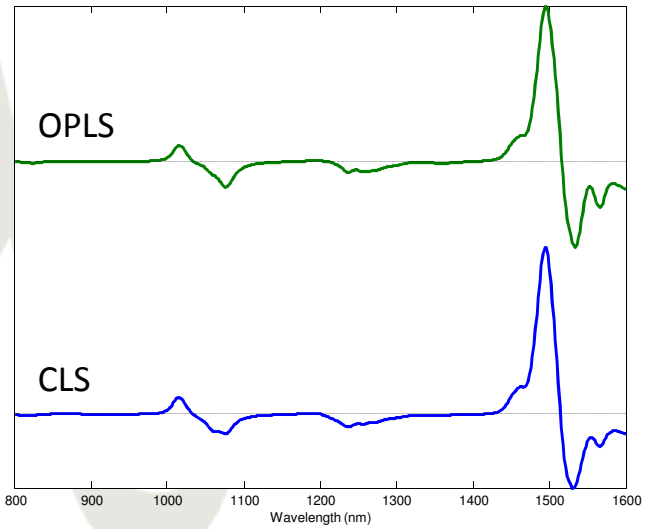
EIGENVECTOR RESEARCH INCORPORATED

Heptane



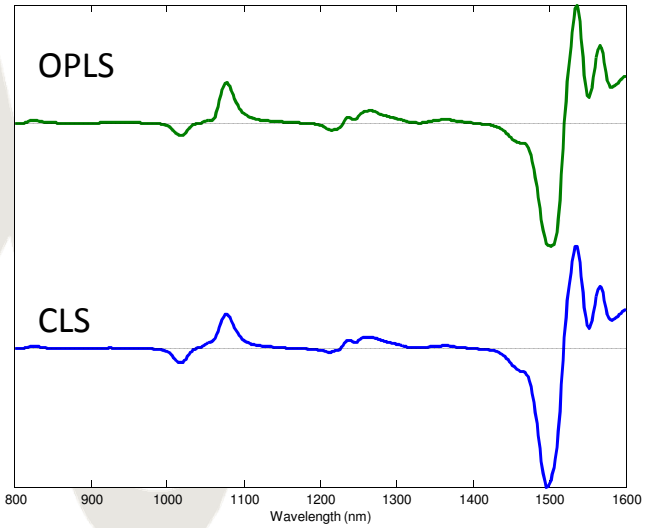
EIGENVECTOR RESEARCH INCORPORATED

Toluene



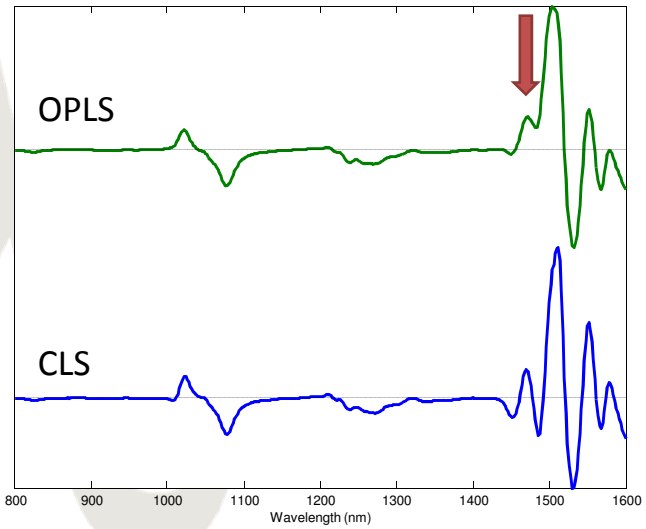
EIGENVECTOR RESEARCH INCORPORATED

Decane



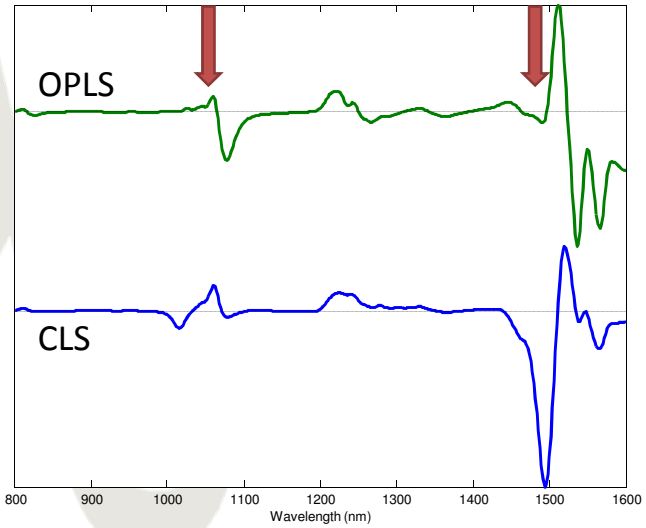
EIGENVECTOR RESEARCH INCORPORATED

Xylene



EIGENVECTOR RESEARCH INCORPORATED

Iso-Octane



EIGENVECTOR RESEARCH INCORPORATED

OPLS vs. CLS?

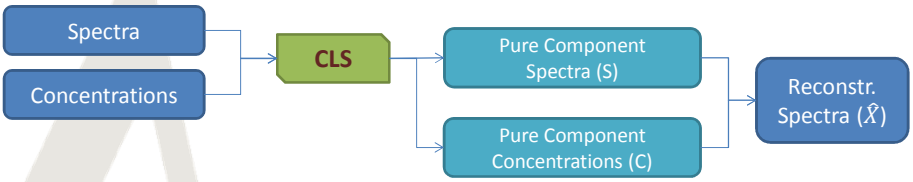
What leads to the difference?

- Differences in noise sensitivity and rotational ambiguity? (i.e. both are correct and equivalent)
- A difference of "opinion"? (i.e. neither is "correct", the real answer is something else)
- Inaccuracy in the OPLS rotation? (i.e. OPLS is "wrong" or requires judicious interpretation)

EIGENVECTOR RESEARCH INCORPORATED

OPLS on Noiseless Data

Is it due to sensitivity to noise and/or a difference of "opinion"?



Calcuate $\hat{X}$=CS

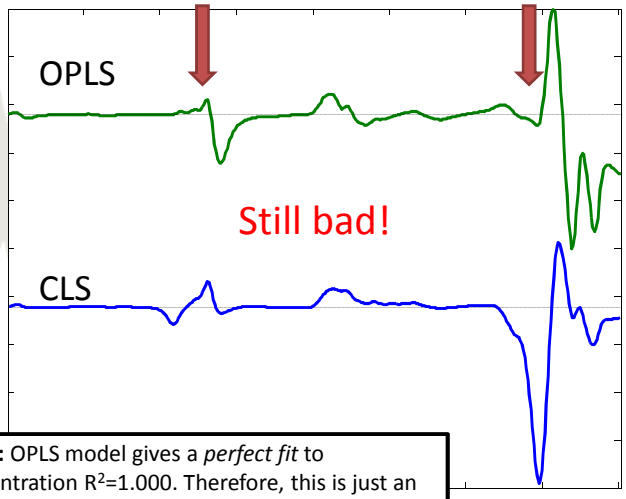
$\hat{X}$ is a rank 5 estimate of X (no noise!)

Use $\hat{X}$ and C in OPLS

This has an exact known answer…

EIGENVECTOR RESEARCH INCORPORATED

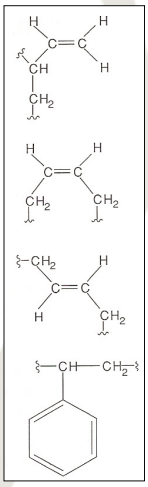
Iso-Octane From *Noiseless* X



NOTE: OPLS model gives a *perfect fit* to concentration R^2=1.000. Therefore, this is just an artifact of the OPLS rotation.

EIGENVECTOR RESEARCH INCORPORATED

Styrene(butadiene) Copolymer



1,2-

cis-

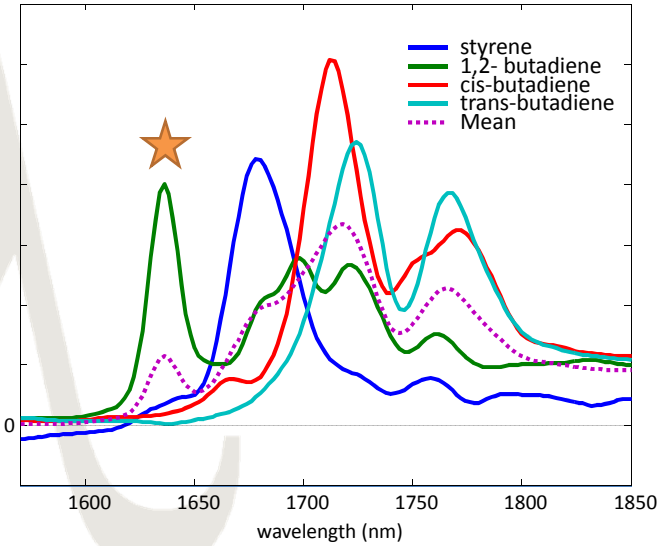
trans-

styrene

- NIR spectra of styrene(butadiene) copolymers
- Different amounts of 4 functional groups. **All 4** are known for **all** samples (by NMR)

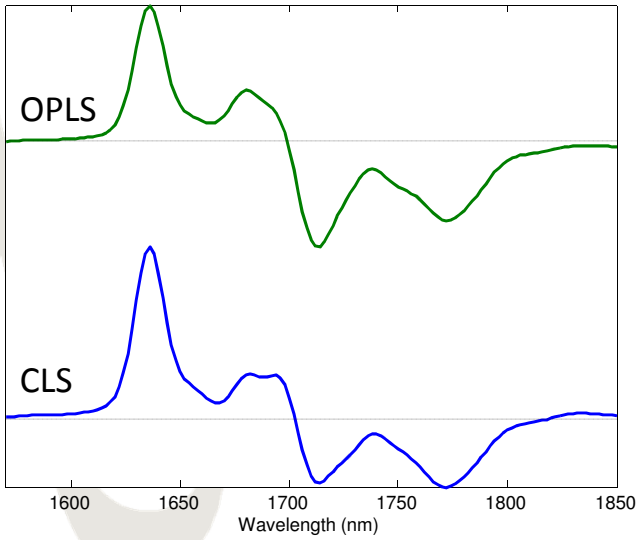
EIGENVECTOR RESEARCH INCORPORATED

CLS-Recovered Pure Component Spectra



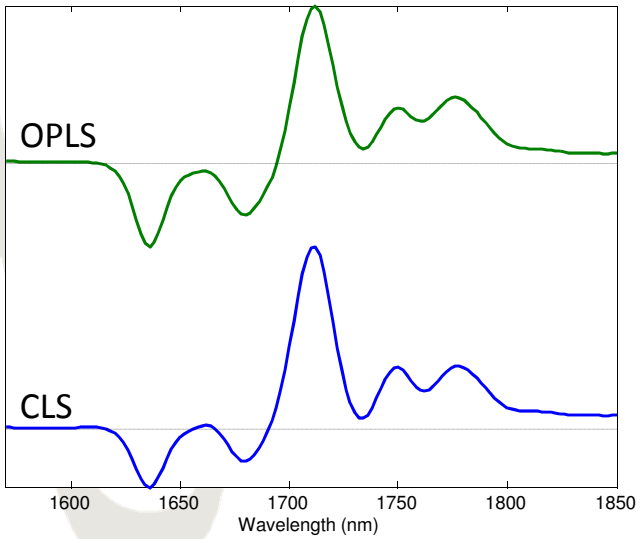
EIGENVECTOR RESEARCH INCORPORATED

1,2-butadiene



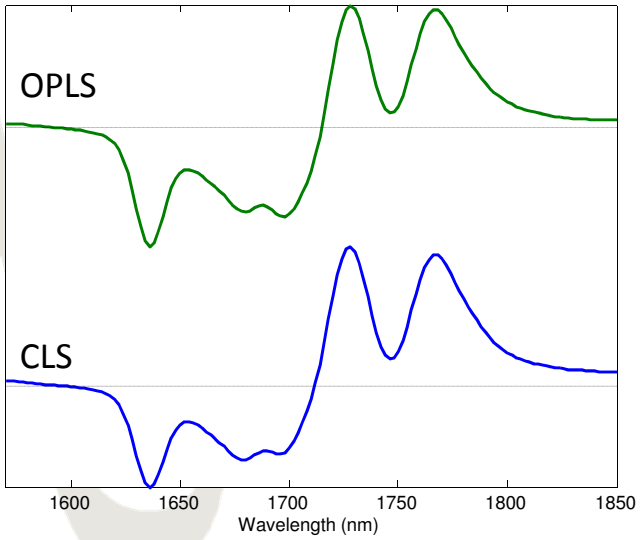
EIGENVECTOR RESEARCH INCORPORATED

cis-butadiene



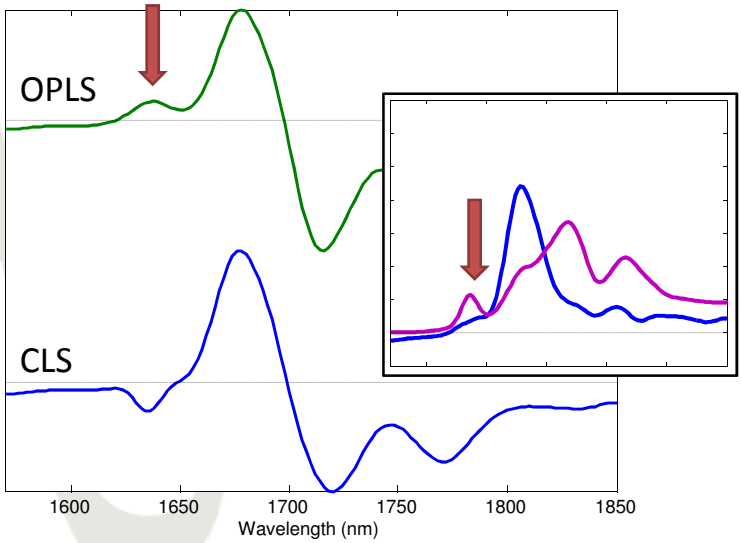
EIGENVECTOR RESEARCH INCORPORATED

trans-butadiene



EIGENVECTOR RESEARCH INCORPORATED

styrene



EIGENVECTOR RESEARCH INCORPORATED

Binary Expression Simulation

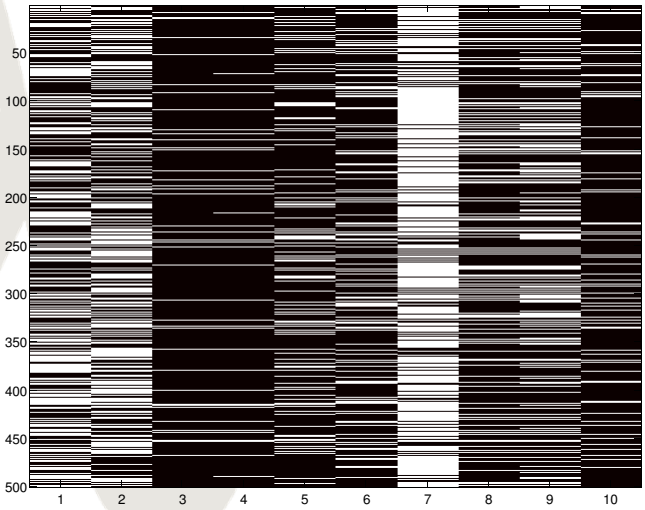
- 10 Expressed "Proteins" (variables), 500 Subjects
- 1 primary effect with loading:

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 resistance)

EIGENVECTOR RESEARCH INCORPORATED

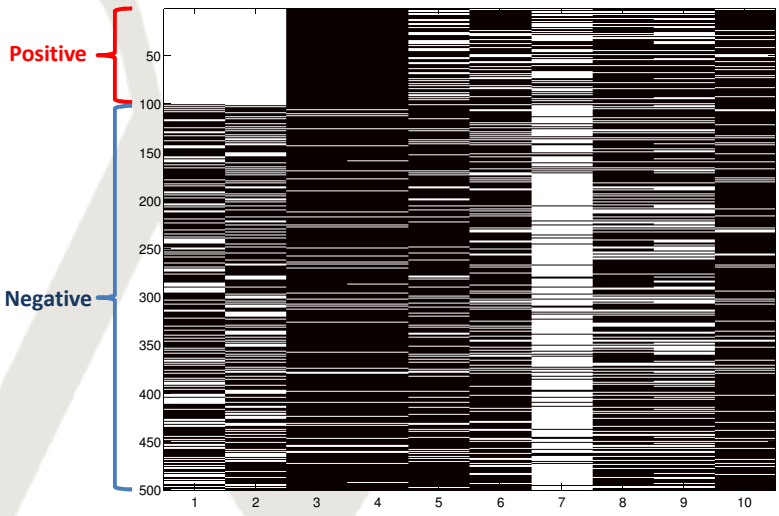
Example Map of Expression



Example Set 1

EIGENVECTOR RESEARCH INCORPORATED

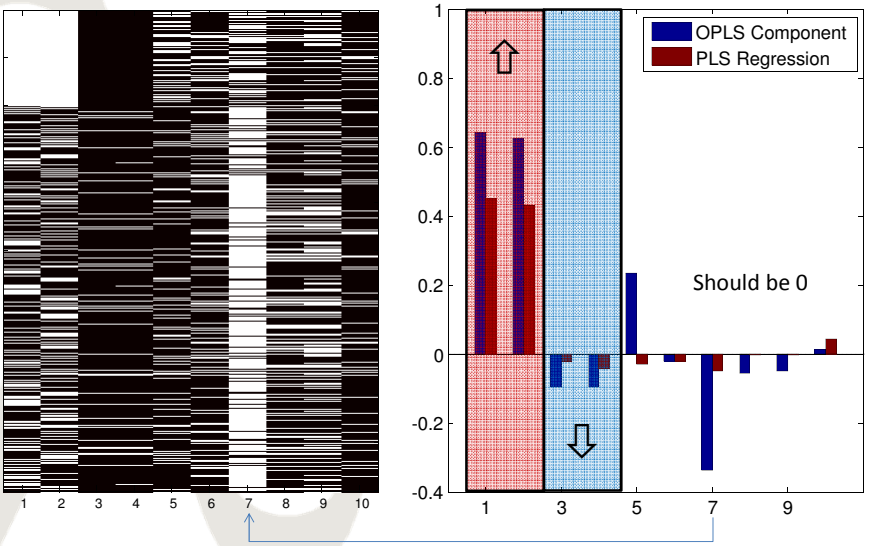
…Sorted by Property of Interest



Example Set 1

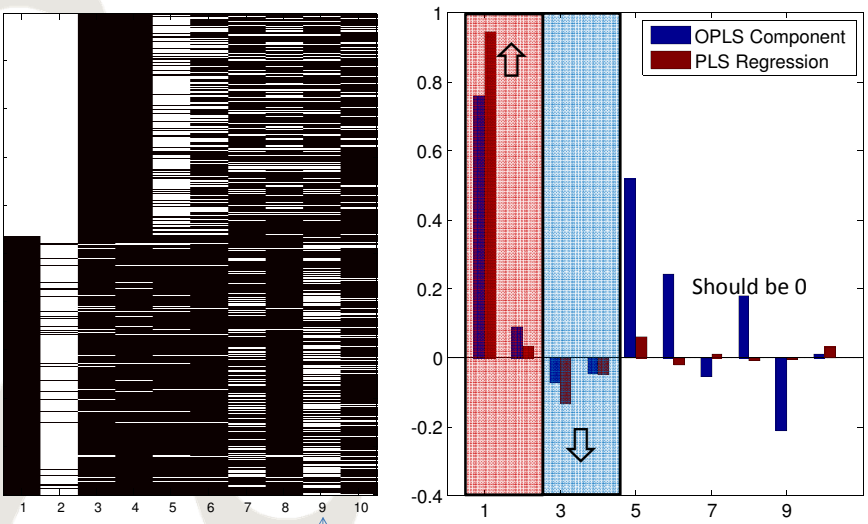
EIGENVECTOR RESEARCH INCORPORATED

Example Set 1



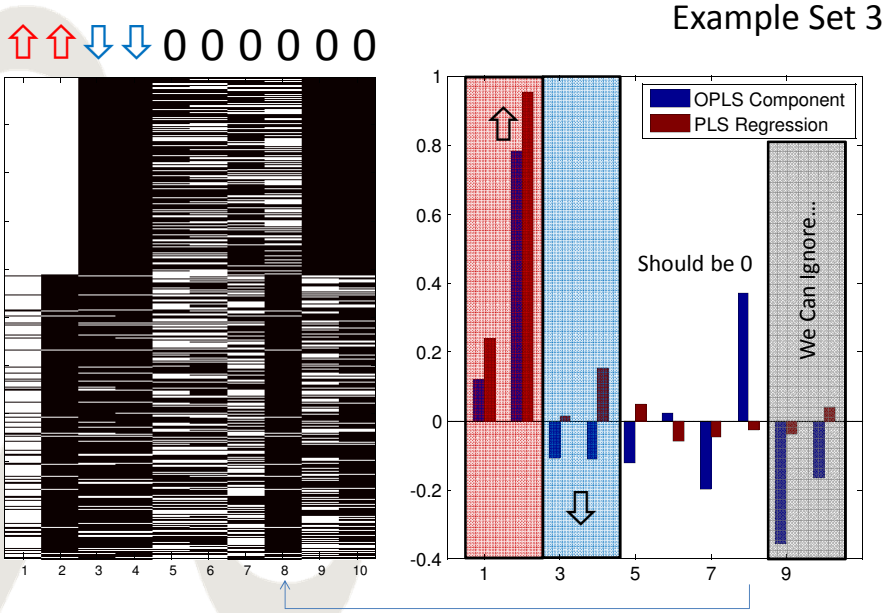
EIGENVECTOR RESEARCH INCORPORATED

Example Set 2



EIGENVECTOR RESEARCH INCORPORATED

Example Set 3



EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- OPLS does simplify regression vectors. It is CLOSER to underlying bilinear response…
- … HOWEVER, result NOT always the same as a first principles model, even after accounting for Mean Centering.
- OPLS recovered component is more sensitive to chance correlation than is regression vector (Problem seen even with 500, 1000, or 2000 samples!)

EIGENVECTOR RESEARCH INCORPORATED