# THINKING OUTSIDE THE *POX*- Alternatives for Stable Model Design

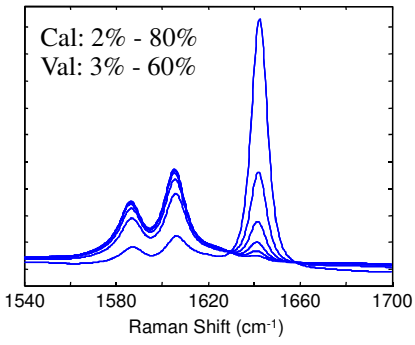
Jeremy M. Shaver, Barry M. Wise, Neal B. Gallagher
Eigenvector Research, Inc.
Manson, WA 98831

- Raman Scattering Data and issues
- Regression & Calibration Transfer Tools
  - GLS - Generalized Least Squares
  - DWPDS - Double Window PDS
  - Full-Ratio PCR
- Trial Calibration Transfers of Raman Data
- Conclusions

EIGENVECTOR RESEARCH INCORPORATED

# Raman Scattering Data

- Narrow bands
- High selectivity
- X-axis variation
- "Pathlength" changes
- Eg. system:
  - Octene (C8) in Toluene
  - 6 Calibration samples, 5 Validation samples
  - 3 Instruments




EIGENVECTOR RESEARCH INCORPORATED

# Regression Tools

- Partial Least Squares (PLS)
  - Standard Inverse Regression with Latent Variables

$$\hat{Y} = Xb$$

  - Preprocessing for all PLS models using MSC and Mean Centering
- Full-Ratio Principal Component Regression (Full-Ratio PCR)
  - Inverse Regression with Principal Components
  - Simultaneous correction for pathlength or throughput differences

$$\hat{Y} = \frac{Xb_1}{Xb_2}$$

EIGENVECTOR RESEARCH INCORPORATED

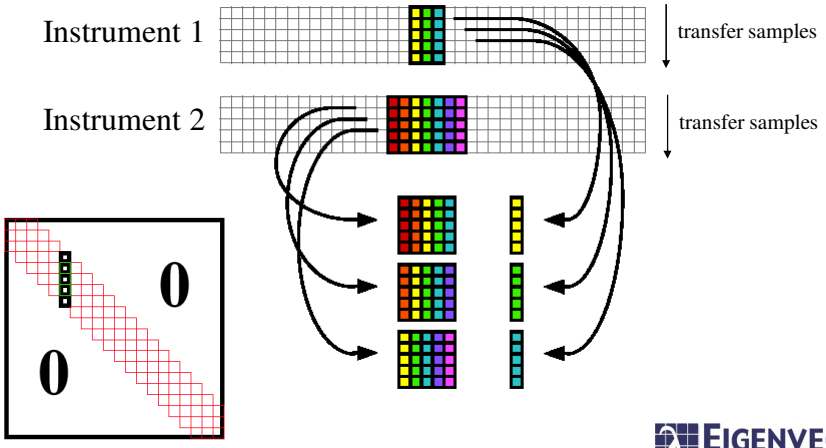
# Instrument Standardization and Calibration Transfer Tools

- Double Window Piecewise Direct Standardization (DWPDS)
  - Make response from Instrument 2 look like Instrument 1
- Generalized Least Squares (GLS)
  - De-weight variables showing difference between Instrument 1 and 2
  - Applied to data from both Instruments 1 and 2.

EIGENVECTOR RESEARCH INCORPORATED

# Double Window Piecewise Direct Standardization (DWPDS)

$$S_1 = S_2 F_b + 1b_s\, T$$




EIGENVECTOR RESEARCH INCORPORATED

# GLS for Calibration Transfer

(1) Calculate the difference between selected (high leverage) transfer samples measured on both instruments.

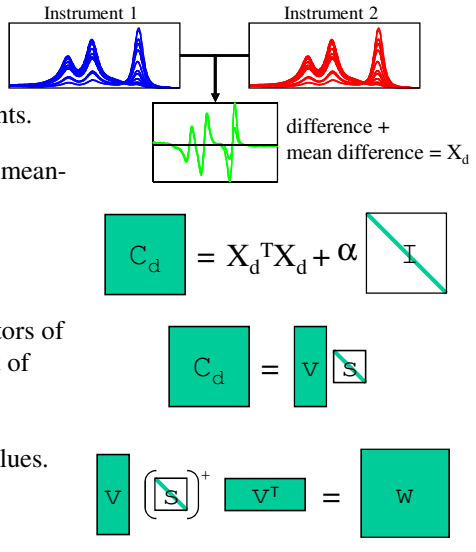
(2) Calculate covariance matrix ($C_d$) of mean-adjusted differences.

(3) Add diagonal (alpha) to $C_d$ matrix.

(4) Calculate eigenvalues and eigenvectors of CV matrix; Select and normalize set of non-zero eigenvalues.

(5) Calculate pseudo inverse of eigenvalues.

(6) Calculate weighting matrix from eigenvalue inverse and eigenvectors.




EIGENVECTOR RESEARCH INCORPORATED

# GLS for Calibration Transfer (2)

- Weighting matrix de-emphasizes features (variables) which are different between the two instruments.


- Original data is weighted using `w` and new model built.

EIGENVECTOR RESEARCH INCORPORATED

# Comparison of Standardization Techniques on NIR Corn Data

**Moisture**

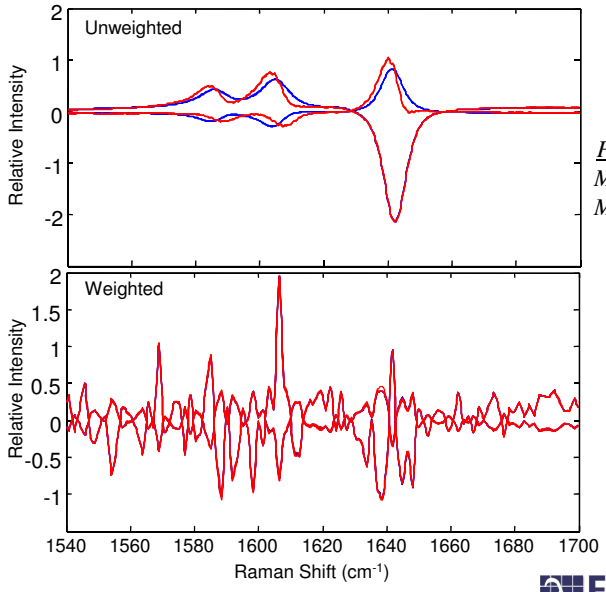
Modeled Instrument / Transfer Method

| Prediction | M5 | MP5 | MP6 |
|---|---|---|---|
| M5 | 0.0187 | 1.4166 | 1.5123 |
| MP5 | 1.1693 | 0.1460 | 0.3547 |
| MP6 | 1.0921 | 0.2849 | 0.1667 |
| **PDS Standardization** | | | |
| M5 | - | 0.3951 | 0.4671 |
| MP5 | 0.2342 | - | 0.1749 |
| MP6 | 0.2068 | 0.1601 | - |
| **GLS Standardization, LVs hand selected** | | | |
| M5 | - | 0.1592 | 0.1908 |
| MP5 | 0.1391 | - | 0.1477 |
| MP6 | 0.1990 | 0.1521 | - |
| **GLS Standardization, best over 5-7 LVs** | | | |
| M5 | - | 0.1545 | 0.1908 |
| MP5 | 0.1248 | - | 0.1239 |
| MP6 | 0.1902 | 0.1177 | - |
| **OSC Standardization, best over all cases,** | | | |
| M5 | - | 0.1630 | 0.1733 |
| MP5 | 0.1945 | - | 0.1580 |
| MP6 | 0.1466 | 0.1320 | - |

Presented by Barry M. Wise, Harald Martens, Martin Høy, Rasmus Bro and Per B. Brockhoff at SSC7 Aug, 2001. 80 corn samples measured on three instruments (M5, MP5 and MP6) Data provided by Mike Blackburn at Cargill.

EIGENVECTOR RESEARCH INCORPORATED

## Calculation of GLS Weightings

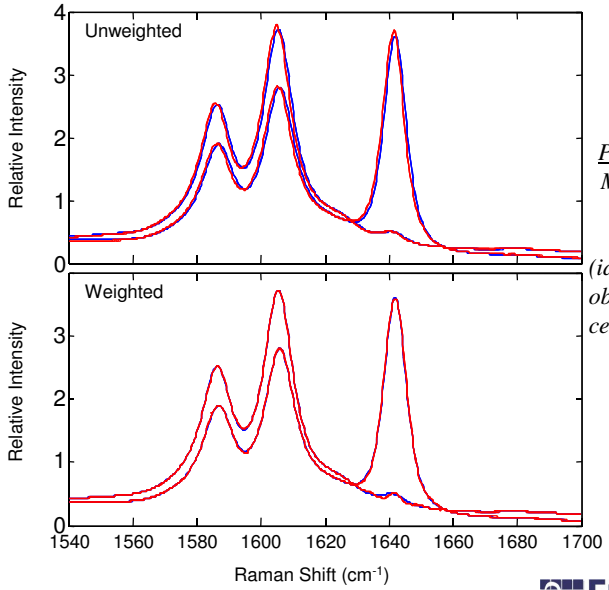



*Preprocessed with:*
*MSC*
*Mean centering*

2 samples; blue = instrument 1, red = instrument 2

EIGENVECTOR RESEARCH INCORPORATED

## Calculation of DWPDS Transform Function




*Preprocessed with:*
*MSC*

*(identical to results obtained with mean centered data)*

2 samples; blue = instrument 1, red = instrument 2

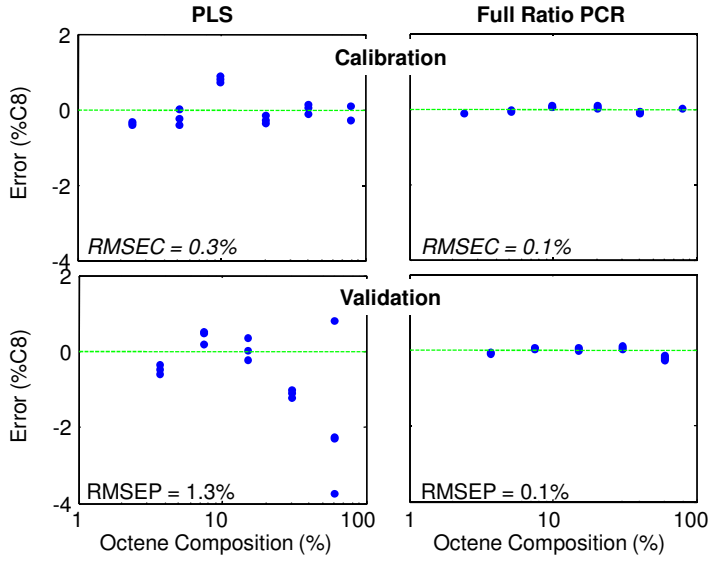
EIGENVECTOR RESEARCH INCORPORATED

# Calibration and Validation on Instrument 1

EIGENVECTOR RESEARCH INCORPORATED

## Instrument 1 - Calibration Residuals

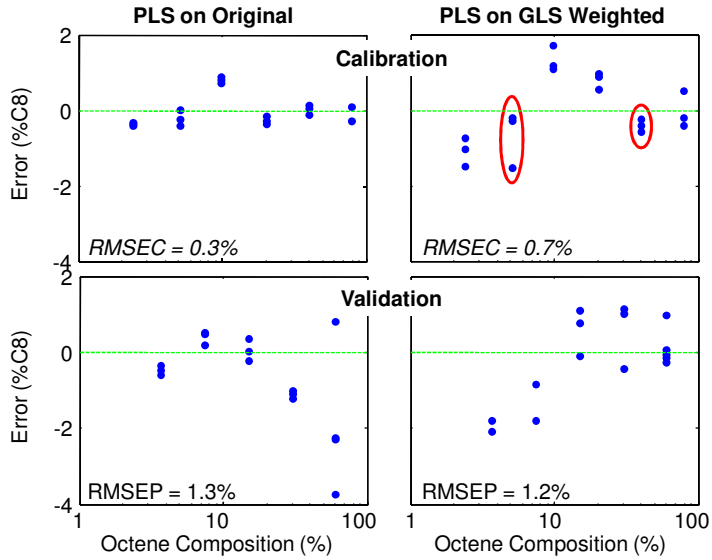
(2 factors)




EIGENVECTOR RESEARCH INCORPORATED

## Instrument 1 - PLS Calibration Residuals

(2 factors) Without and With GLS Weighting




O = samples used for weighting calculation

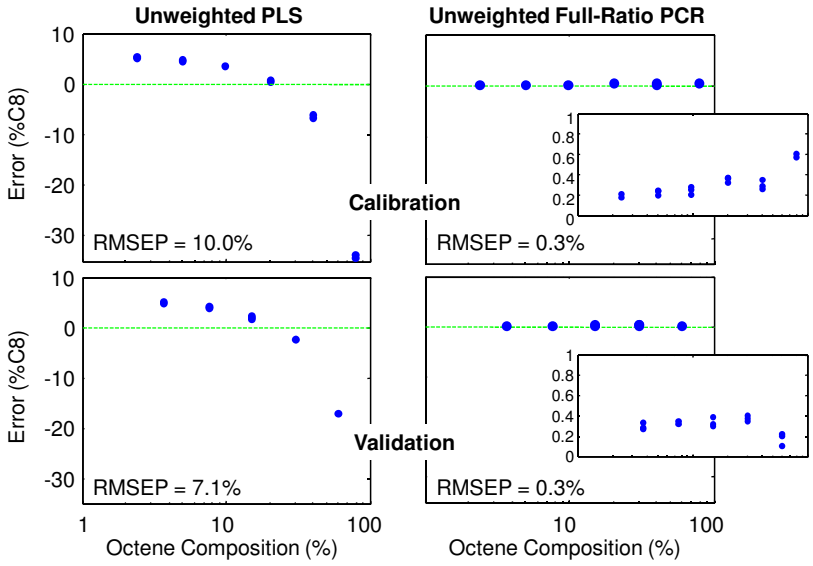
EIGENVECTOR RESEARCH INCORPORATED

## Prediction on Instrument 2

EIGENVECTOR RESEARCH INCORPORATED

## Instrument 2 - Prediction Residuals

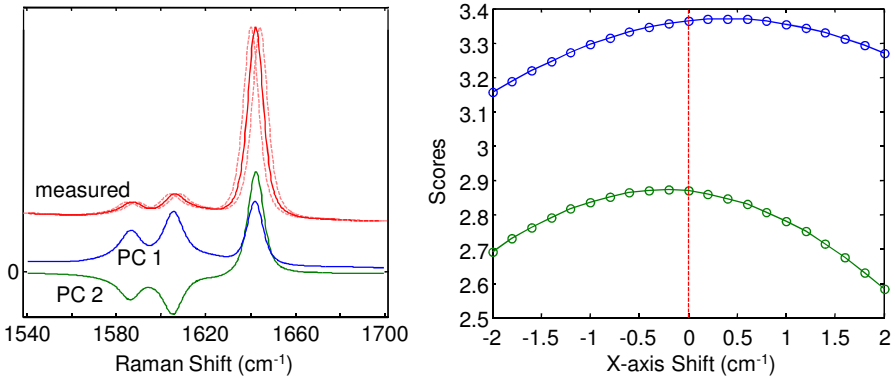
(UNTRANSFERED RAW DATA - 2 factors)



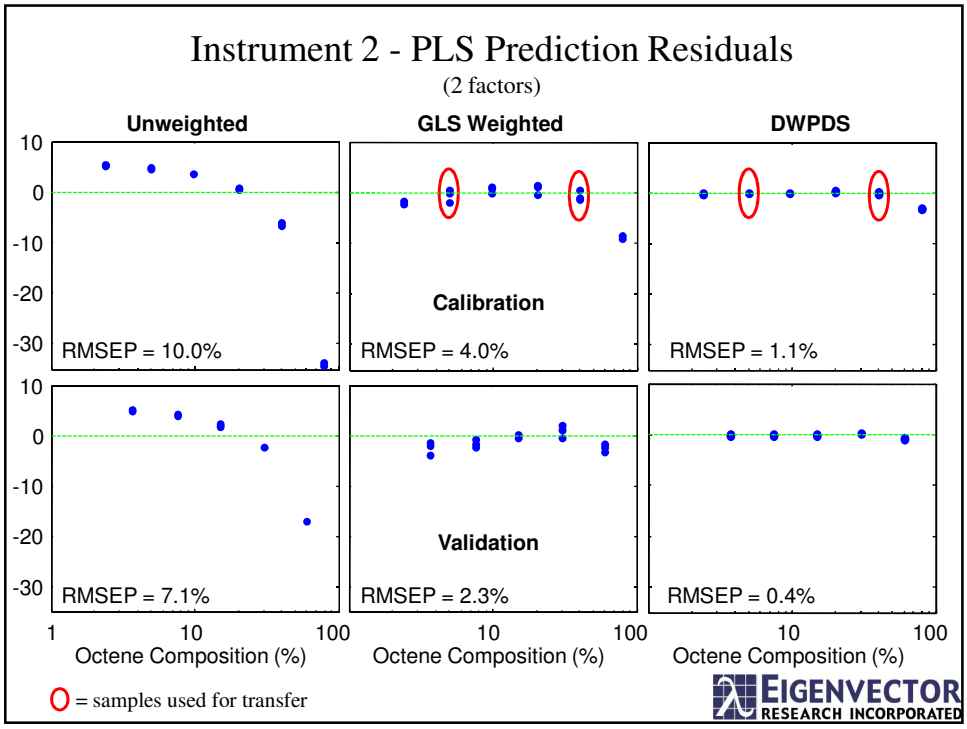

EIGENVECTOR RESEARCH INCORPORATED

## Why Does Full-Ratio Work for Calibration Transfer?

***Relative*** scores change with small shift ≈ a scaling effect!




EIGENVECTOR RESEARCH INCORPORATED




# Prediction Results

(MSC)

| | | PLS | | | FullRatio-PCR | |
|---|---|---|---|---|---|---|
| | None | GLS | DWPDS | None | DWPDS |
| **Inst 1** | | | | | | |
| Cal | *0.3%* | *0.7%* | | *0.1%* | |
| **Val** | **1.3%** | **1.2%** | | **0.1%** | |
| **Inst 2** | | | | | | |
| Cal | 10.0% | 4.0% | 1.1% | 0.3% | 0.2% |
| **Val** | **7.1%** | **2.3%** | **0.4%** | **0.3%** | **0.2%** |
| **Inst 3** | | | | | | |
| Cal | 0.8% | 1.6% | 0.8% | 0.2% | 0.1% |
| **Val** | **0.8%** | **0.9%** | **0.6%** | **0.2%** | **0.1%** |

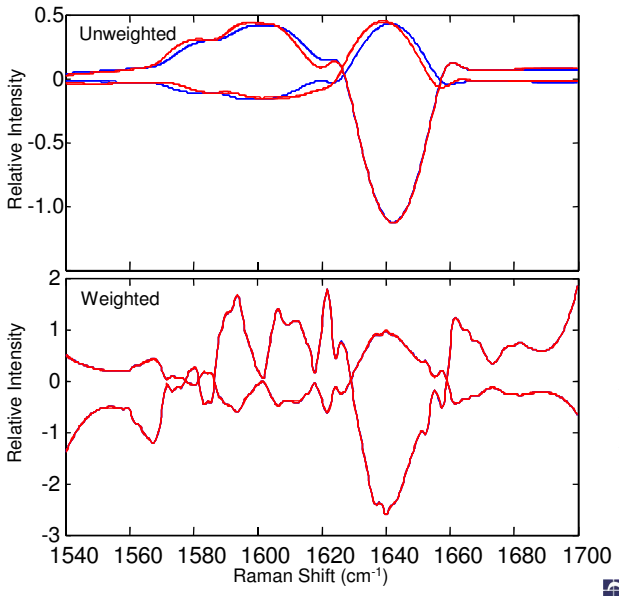
2.5x-3x improvement

10x-18x improvement

23x-33x improvement

EIGENVECTOR RESEARCH INCORPORATED

## Calculation of Weightings (On Blurred Spectra)

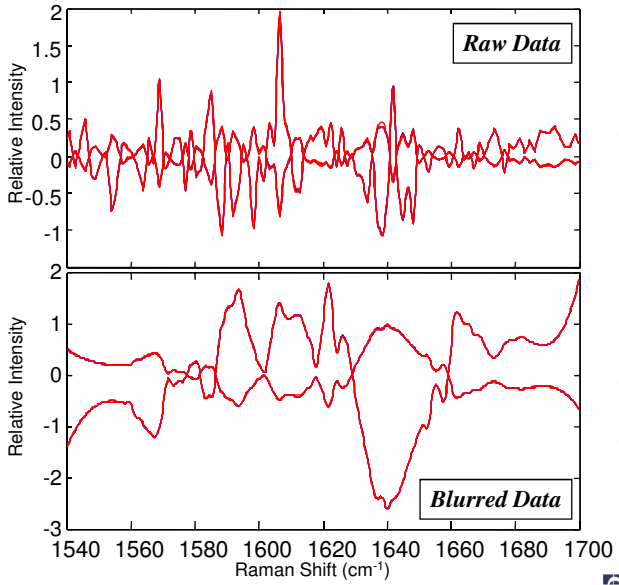



*Preprocessed with:*
*MSC*
***36 cm$^{-1}$ SG Smooth***
*Mean centering*

2 samples; blue = instrument 1, red = instrument 2

EIGENVECTOR RESEARCH INCORPORATED

## GLS Weighted Spectra




*Preprocessed with:*
*MSC*
*Mean centering*

*Preprocessed with:*
*MSC*
***36 cm$^{-1}$ SG Smooth***
*Mean centering*

2 samples; blue = instrument 1, red = instrument 2

EIGENVECTOR RESEARCH INCORPORATED

# Prediction Results

(MSC + *36 cm$^{-1}$ smooth*)

| | PLS* | | | FullRatio-PCR | |
|---|---|---|---|---|---|
| | None | GLS | DWPDS | None | DWPDS |
| **Inst 1** | | | | | |
| Cal | 0.3% | 1.0% | | <0.1% | |
| **Val** | **1.0%** | **2.0%** | | **0.1%** | |
| **Inst 2** | | | | | |
| Cal | 9.0% | 1.5% | 0.7% | 0.2% | 0.2% |
| **Val** | **6.4%** | **1.6%** | **0.3%** | **0.2%** | **0.1%** |

4x-6x improvement

13x-21x improvement

32x-45x improvement

EIGENVECTOR RESEARCH INCORPORATED

# Conclusions

GLS

- Useful even with very few transfer samples (depends on complexity of system and instrument difference)
- Less effective when net analyte signal will be badly reduced by weighting (e.g. very narrow bands, large shifts)

DWPDS

- Very good results - window sizes should be selected to always include signal

Full Ratio PCR

- Excellent results ?!? - but can not be combined with mean-centered data.
- Has problems with “invisible” species (no spectral signature)

EIGENVECTOR RESEARCH INCORPORATED

# Acknowledgements