Generalized Least Squares for Calibration Transfer

Barry M. Wise, Harald Martens
and Martin Høy
Eigenvector Research, Inc.
Manson, WA

Outline

- The calibration transfer problem
  - Instrument differences, drift, environment changes
  - Pseudo gasoline
  - Corn
- Generalized Least Squares preprocessing
- PDS and OSC
- Comparison of results
- Conclusions

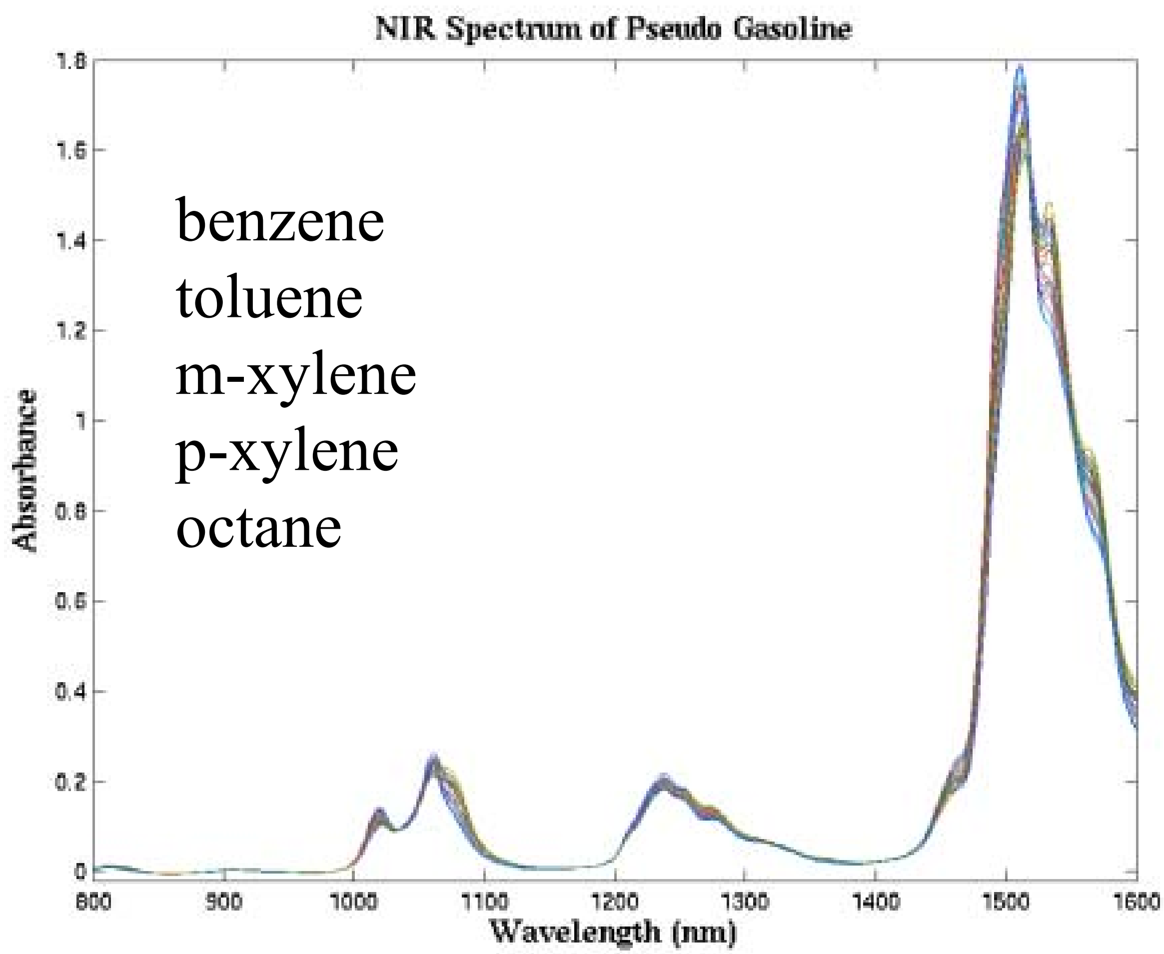
EIGENVECTOR RESEARCH INCORPORATED

Reasons for Calibration Transfer

- No two instruments identical
  - Some calibrations depend on very small changes in data
- Single instruments often drift
  - Aging parts, dirt
  - Temperature
  - *Standardization*
- New interferences in samples

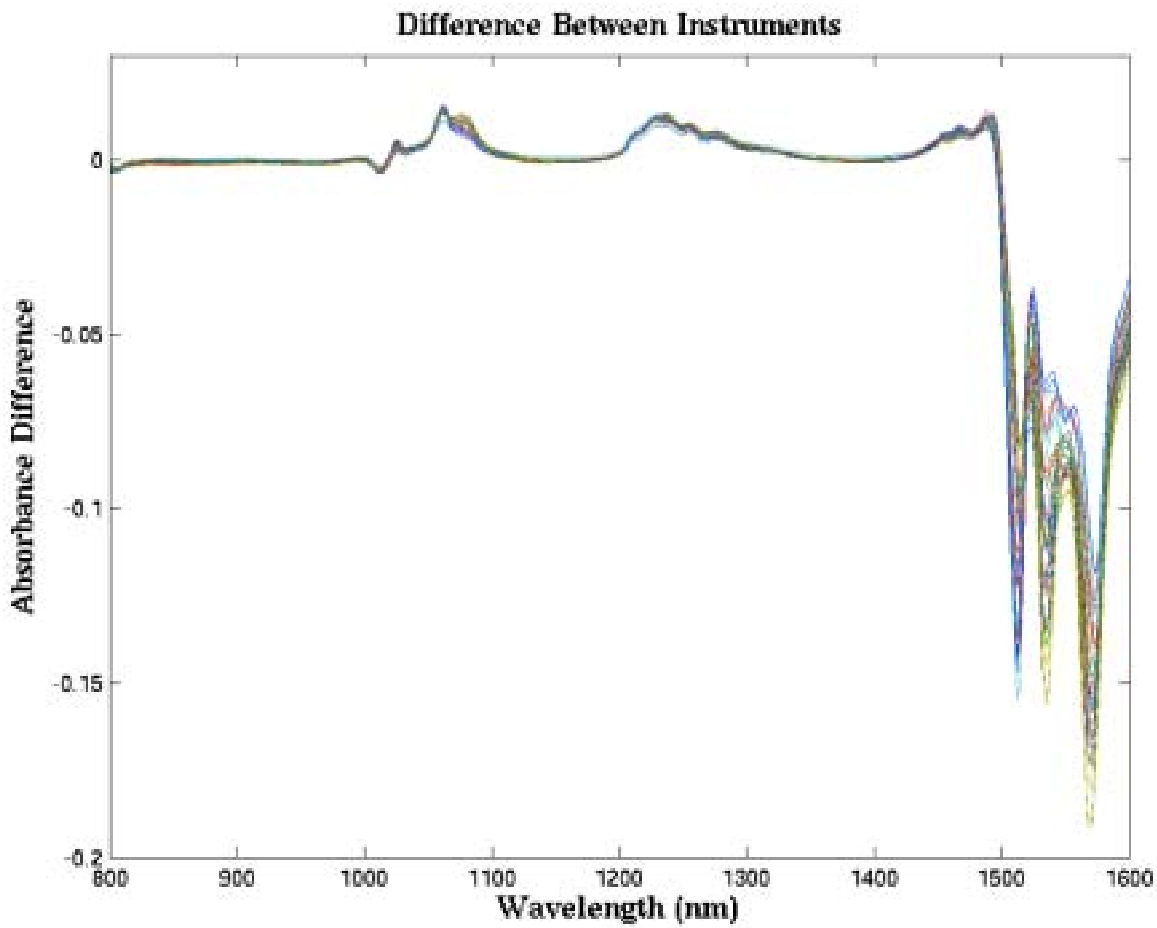
EIGENVECTOR RESEARCH INCORPORATED

Pseudo Gasoline Data



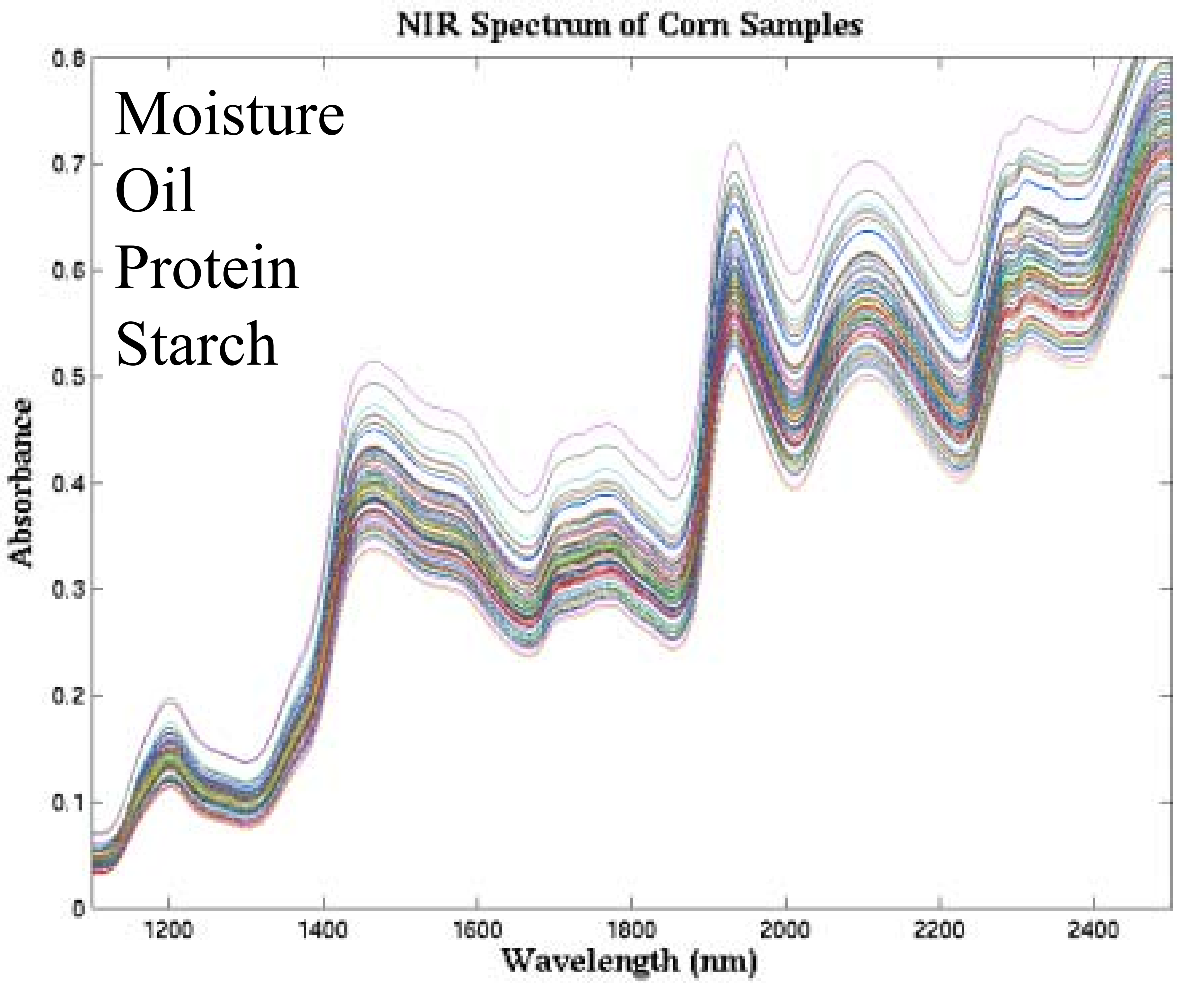
EIGENVECTOR RESEARCH INCORPORATED

Difference Between Instruments

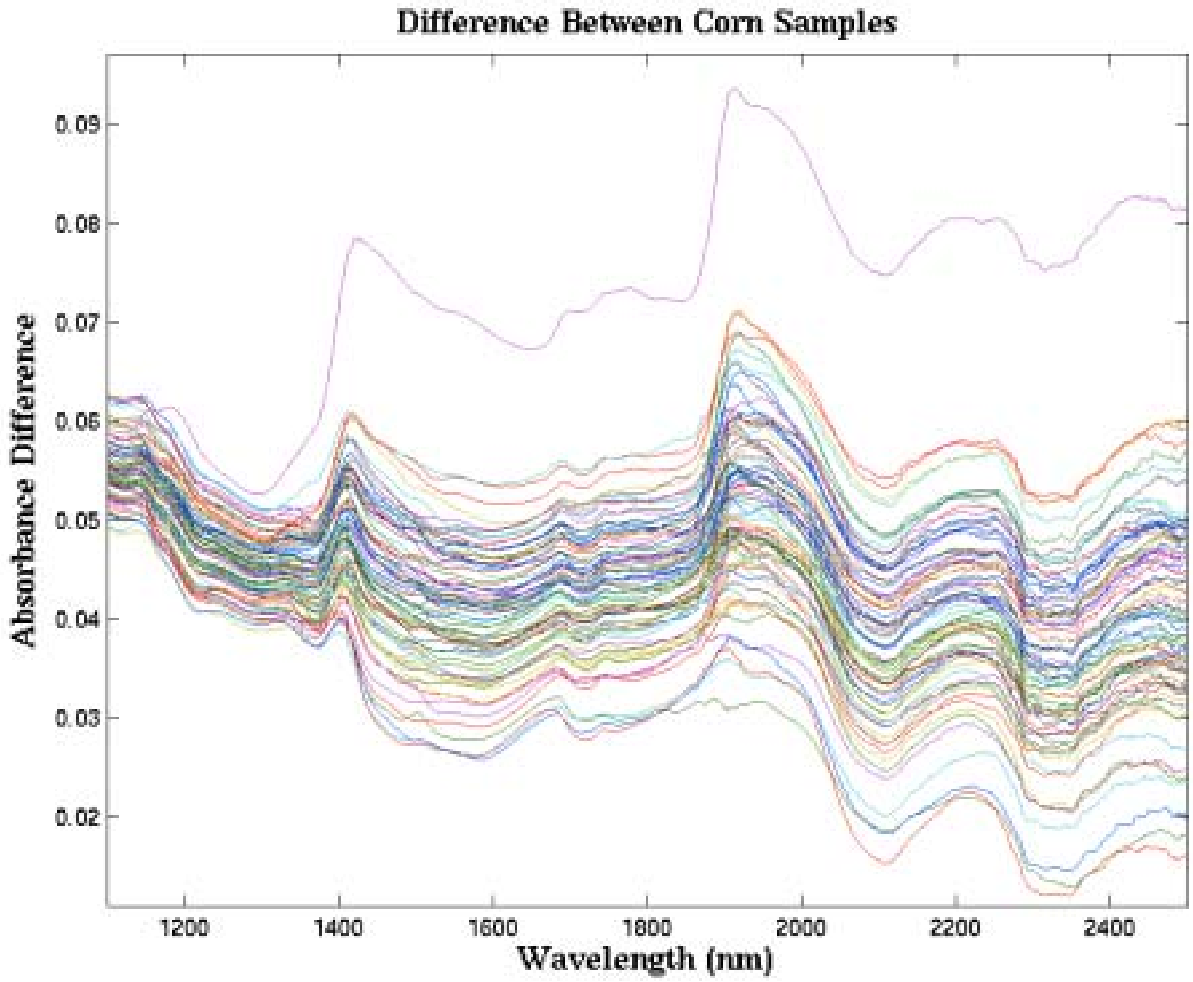


EIGENVECTOR RESEARCH INCORPORATED

Corn Data


NIR Spectrum of Corn Samples
Moisture
Oil
Protein
Starch
Absorbance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1200
1400
1600
1800
2000
2200
2400
Wavelength (nm)

Difference Between Instruments



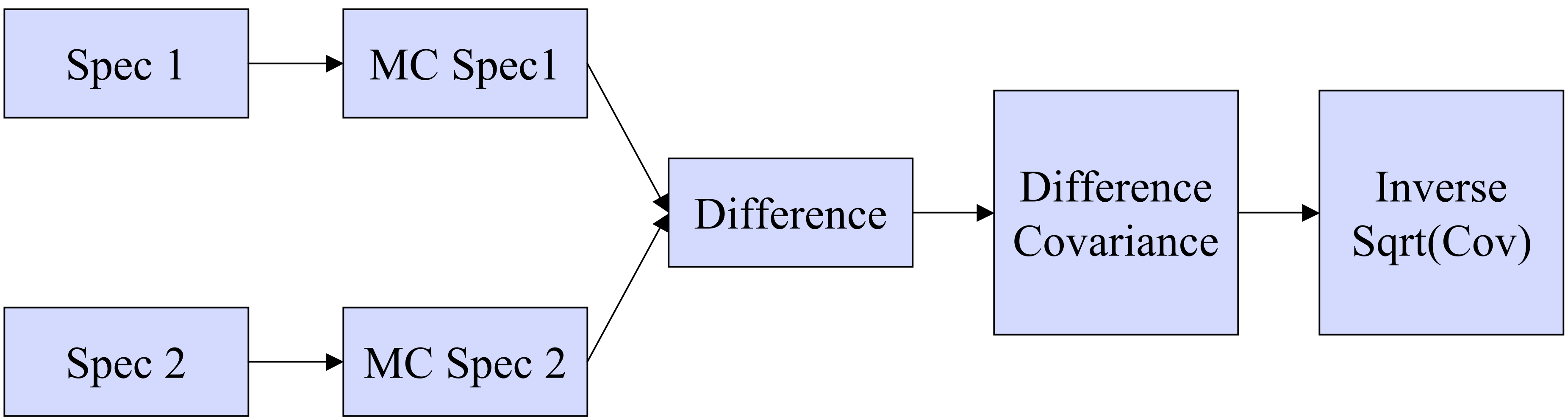
EIGENVECTOR RESEARCH INCORPORATED

Selection of Transfer Samples

- Transfer samples should
  - be "high leverage"
  - span the space of differences
- Several ways to choose
  - Hand select (based on PC scores, etc.)
  - Find high leverage in PCA
  - Find high leverage based on calibration model

EIGENVECTOR RESEARCH INCORPORATED

Development of GLS Weighting Matrix



EIGENVECTOR RESEARCH INCORPORATED

Difference Covariance

$$\mathbf{X}_d = (\mathbf{X}_{1,tr} - \bar{\mathbf{x}}_{1,tr}) - (\mathbf{X}_{2,tr} - \bar{\mathbf{x}}_{2,tr})$$

$$\mathbf{C} = \frac{\mathbf{X}_d^{\mathrm{T}}\mathbf{X}_d}{N-1}$$

Covariance to Weighting Matrix

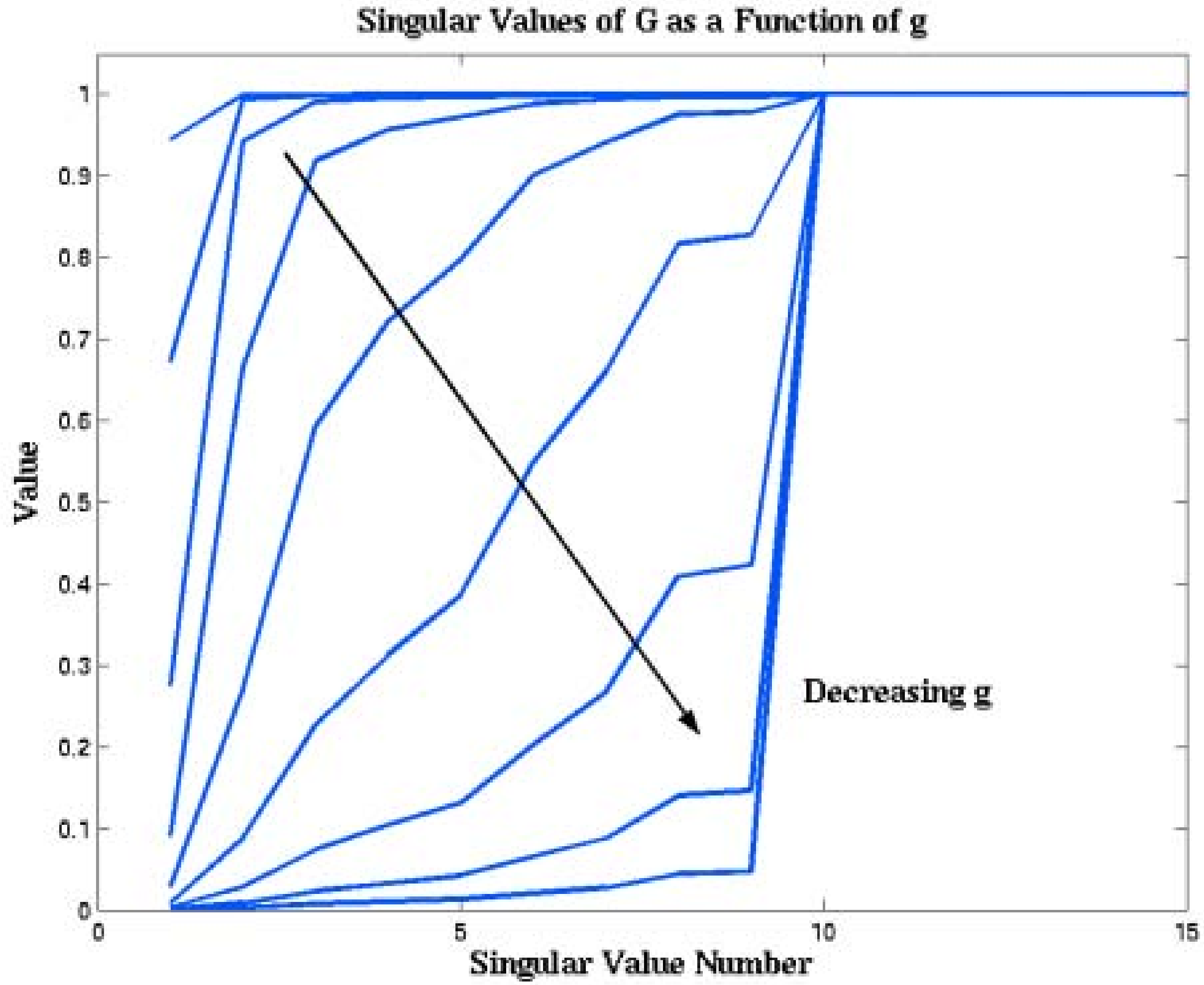
$$\mathbf{C} = \mathbf{V}\mathbf{S}^{2}\mathbf{V}^{\mathrm{T}}$$

$$\mathbf{G} = \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^{\mathrm{T}}$$

$$d_{i,i}^{-1} = \frac{1}{\sqrt{\dfrac{s_{i,i}^{2}}{g^{2}} + 1}}$$

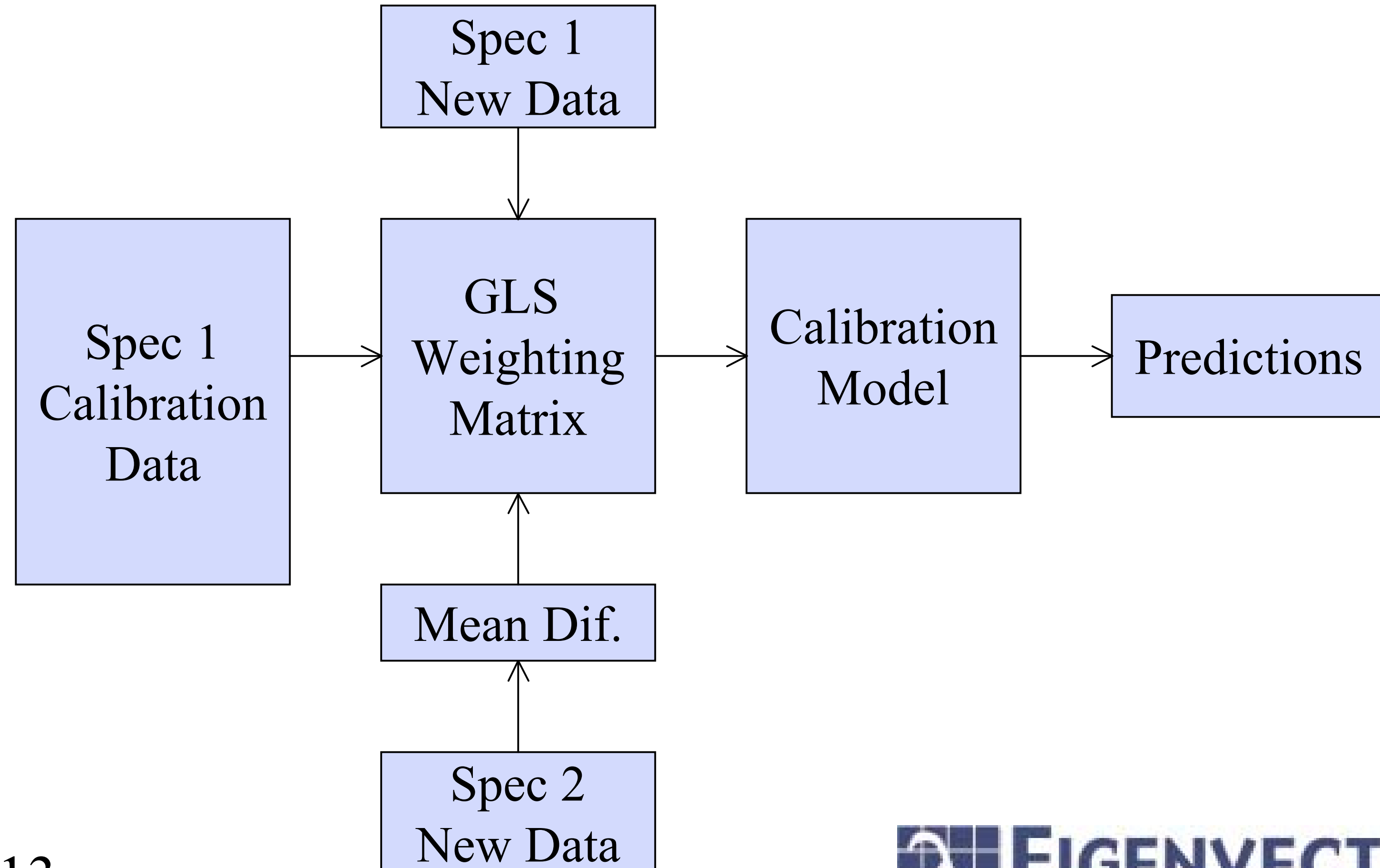
EIGENVECTOR RESEARCH INCORPORATED

Effect of Parameter g

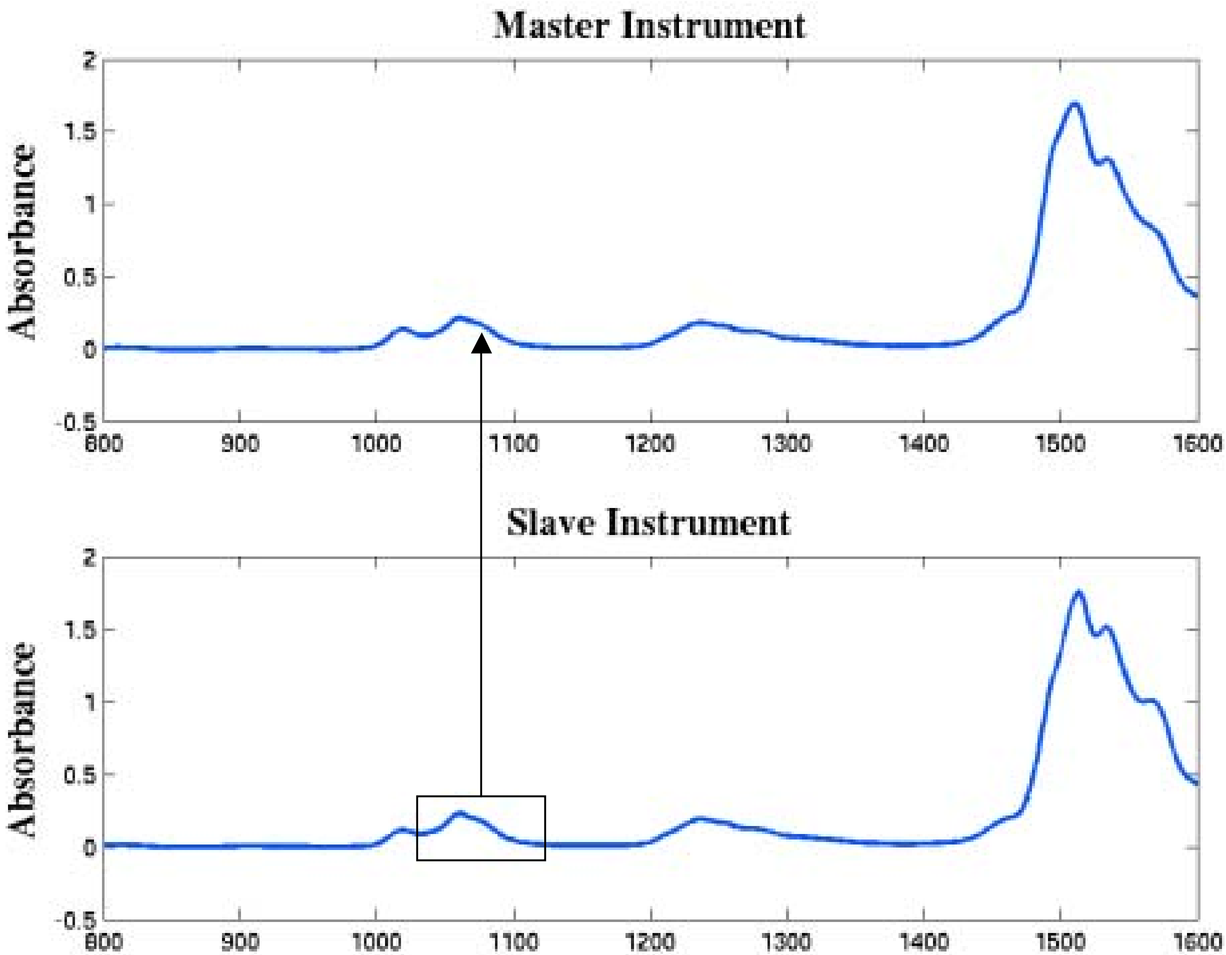


EIGENVECTOR RESEARCH INCORPORATED

Application of GLS Weighting Matrix



Piece-wise Direct Standardization

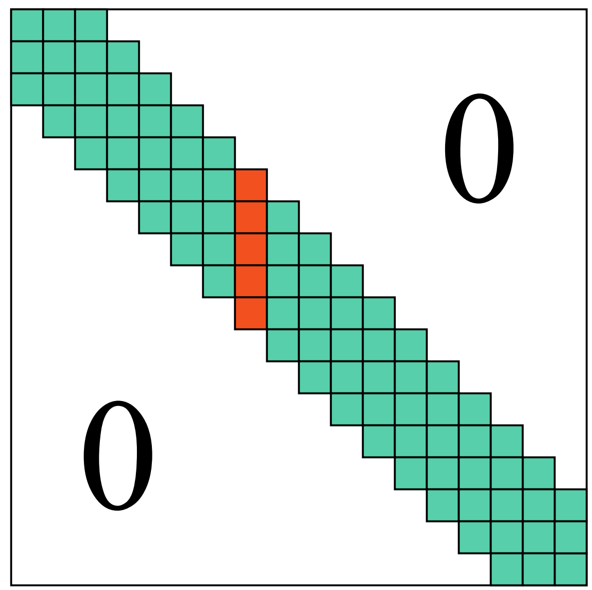


EIGENVECTOR RESEARCH INCORPORATED

PDS Model

$$\mathbf{X}_1 = \mathbf{X}_2\mathbf{F} + \mathbf{1b}_{2-1}$$

F =

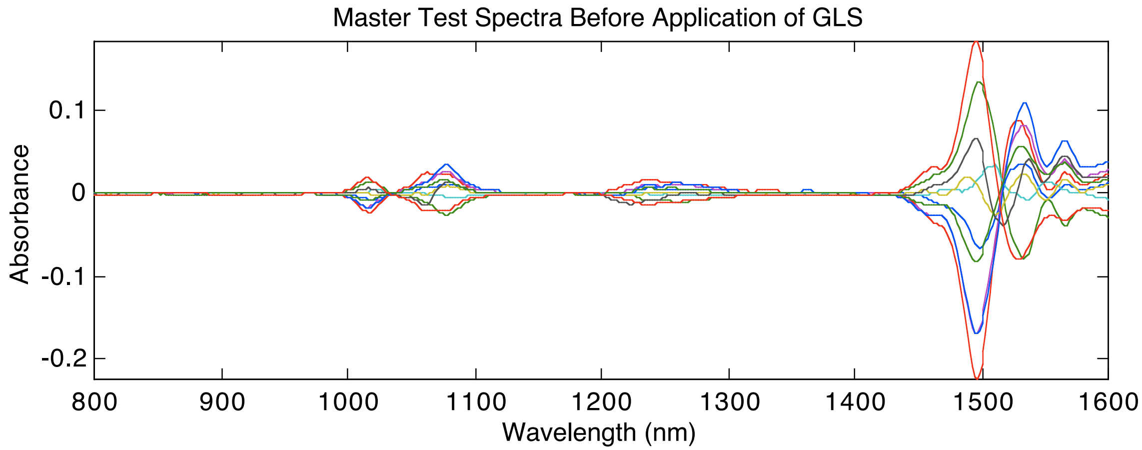


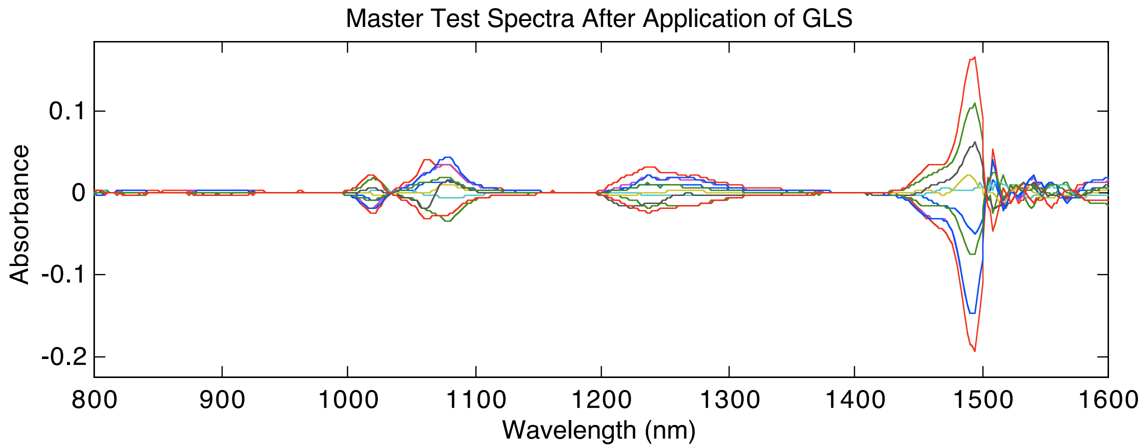
EIGENVECTOR RESEARCH INCORPORATED

Orthogonal Signal Correction

- Determine factor which describes large amounts of variance in **X** while being orthogonal to **Y**
- Deflate **X**
- Build PLS model that predicts scores of deflation factor
- Use PLS model to estimate amount of factor to remove from new **X**

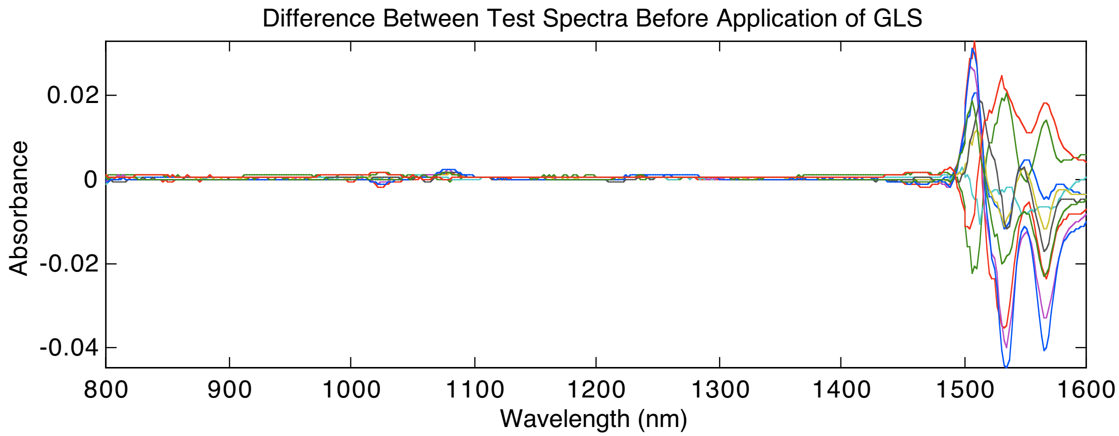
Pseudo Gasoline Master Before and After GLS

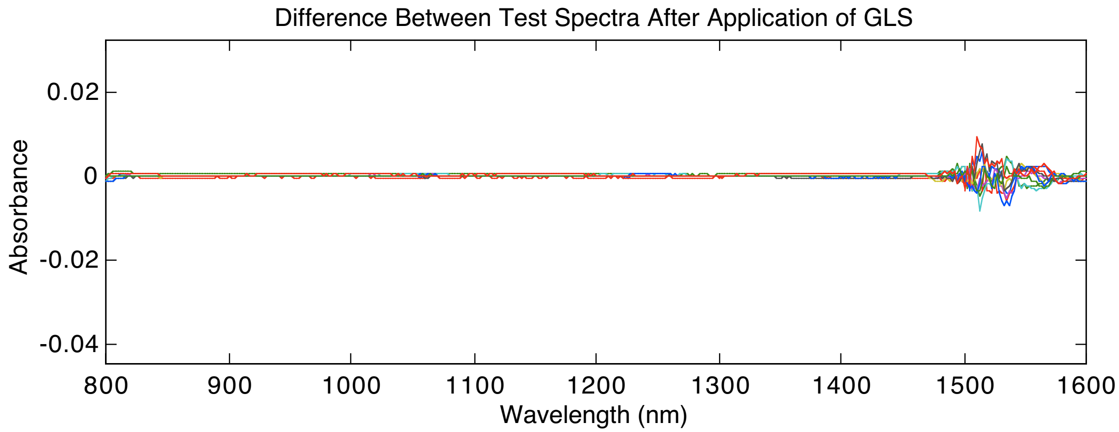




EIGENVECTOR RESEARCH INCORPORATED

Pseudo Gasoline Difference Before and After GLS





⇨movie

EIGENVECTOR RESEARCH INCORPORATED

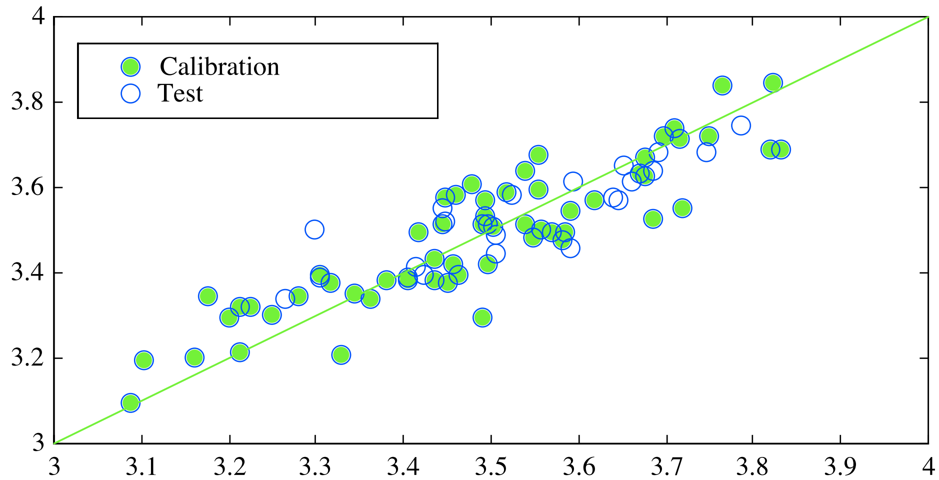
Comparison of Methods on Corn Data

- Available data
  - 80 samples split 60/20
  - 3 instruments
  - 4 analytes
- 5 Transfer samples selected
  - Based on model inverse for PDS
  - Based on PCA leverage for OSC, GLS
- Tested all 3 methods on all combinations of instrument and analyte

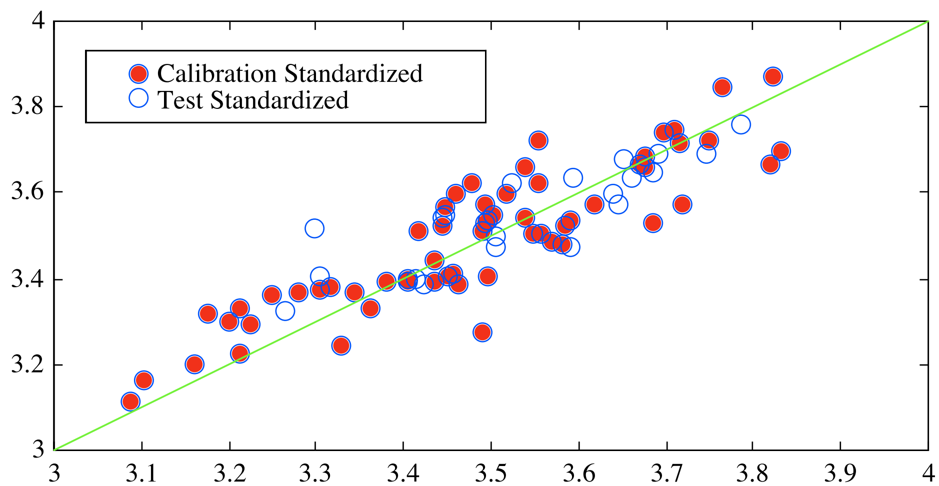
Issues with Meta-parameters

- GLS has only one parameter, g
- PDS
  - Window width
  - Parameters for sub models (LVs or tolerance)
- OSC
  - Number of OSC LVs
  - Tolerance of initial iterations
  - Tolerance on reconstruction
- Number of LVs in PLS calibration models
- *Try to shown each technique in best light!*

Typical Calibration and Test Data



Standardizing MP5 to M5 for Corn moisture



EIGENVECTOR RESEARCH INCORPORATED

Results from Corn Data

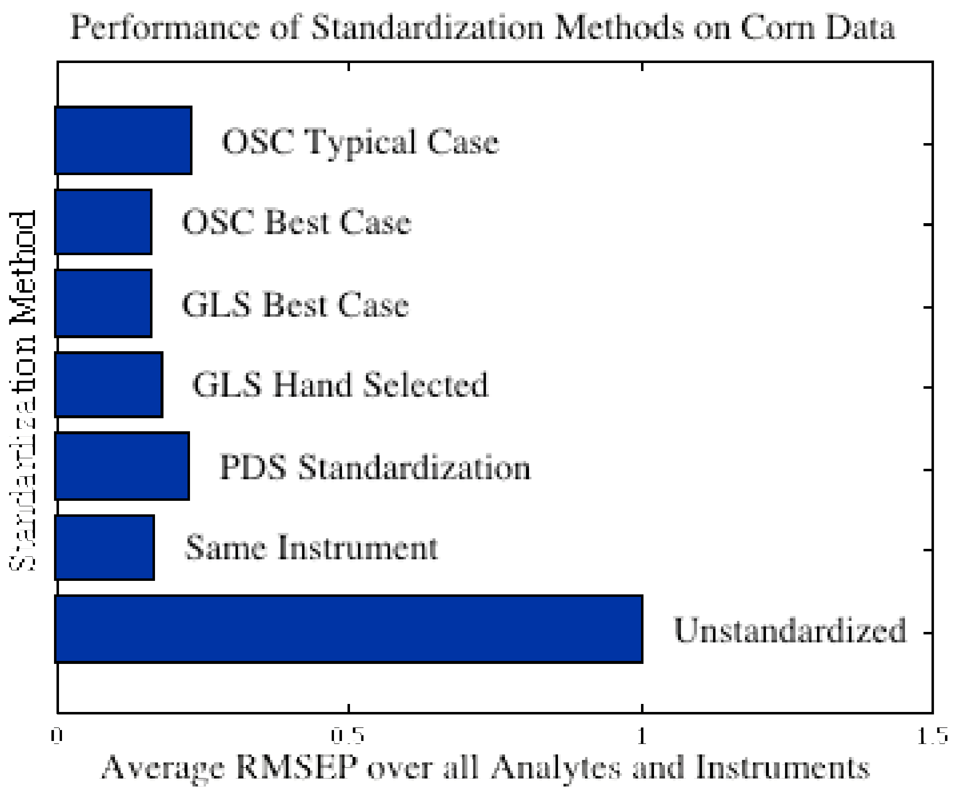
Prediction Instrument

Model Instrument

Preds	Moisture M5	Moisture MP5	Moisture MP6	Oil M5	Oil MP5	Oil MP6	Protein M5	Protein MP5	Protein MP6	Starch M5	Starch MP5	Starch MP6	Average
M5	0.0187	1.4166	1.5123	0.0361	0.1274	0.1568	0.1302	1.2685	1.3241	0.2077	2.0949	1.6601	
MP5	1.1693	0.1460	0.3547	0.2751	0.0885	0.1516	1.2719	0.1720	0.2782	3.5674	0.4091	0.6119	1.0052
MP6	1.0921	0.2849	0.1667	0.3148	0.1926	0.0819	0.8982	0.2403	0.1876	3.1865	0.6754	0.4031	0.1706
PDS Standardization													
M5	–	0.3951	0.4671	–	0.0932	0.0755	–	0.1699	0.1849	–	0.3362	0.3710	
MP5	0.2342	–	0.1749	0.0876	–	0.0944	0.1401	–	0.1880	0.3455	–	0.3972	0.2289
MP6	0.2068	0.1601	–	0.0920	0.1035	–	0.1553	0.1770	–	0.4147	0.4290	–	
GLS Standardization, LVs hand selected													
M5	–	0.1592	0.1908	–	0.0859	0.0952	–	0.1531	0.1679	–	0.3314	0.3420	
MP5	0.1391	–	0.1477	0.0479	–	0.0770	0.1722	–	0.2110	0.2830	–	0.4381	0.1831
MP6	0.1990	0.1521		0.0603	0.0816		0.1687	0.1570		0.1873	0.3473		
GLS Standardization, best over 5-8 LVs													
M5	–	0.1545	0.1897	–	0.0688	0.0783	–	0.1485	0.1602	–	0.3039	0.3350	
MP5	0.1248		0.1258	0.0479		0.0696	0.1448		0.1721	0.2405		0.3709	0.1662
MP6	0.1902	0.1177	–	0.0590	0.0753	–	0.1358	0.1570	–	0.1873	0.3316	–	
OSC Standardization, best over all cases, 1-3 OSC, 3-8 LVs													
M5	–	0.1630	0.1733	–	0.0816	0.0710	–	0.1433	0.1502	–	0.3002	0.3293	
MP5	0.1945		0.1580	0.0710		0.0739	0.1394		0.1988	0.2640		0.4259	0.1637
MP6	0.1466	0.1320		0.0607	0.0686		0.1568	0.1449		0.2253	0.3744		
OSC Standardization, best single case, 3 OSC 5 LVs													
M5		0.2218	0.2611		0.0835	0.0742		0.1588	0.1502		0.3250	0.3515	
MP5	0.3097	–	0.2176	0.0830	–	0.0834	0.1601	–	0.2388	0.4206	–	0.4506	
MP6	0.3299	0.1379	–	0.0826	0.1157	–	0.1684	0.2154	–	0.5119	0.4363	–	

EIGENVECTOR RESEARCH INCORPORATED

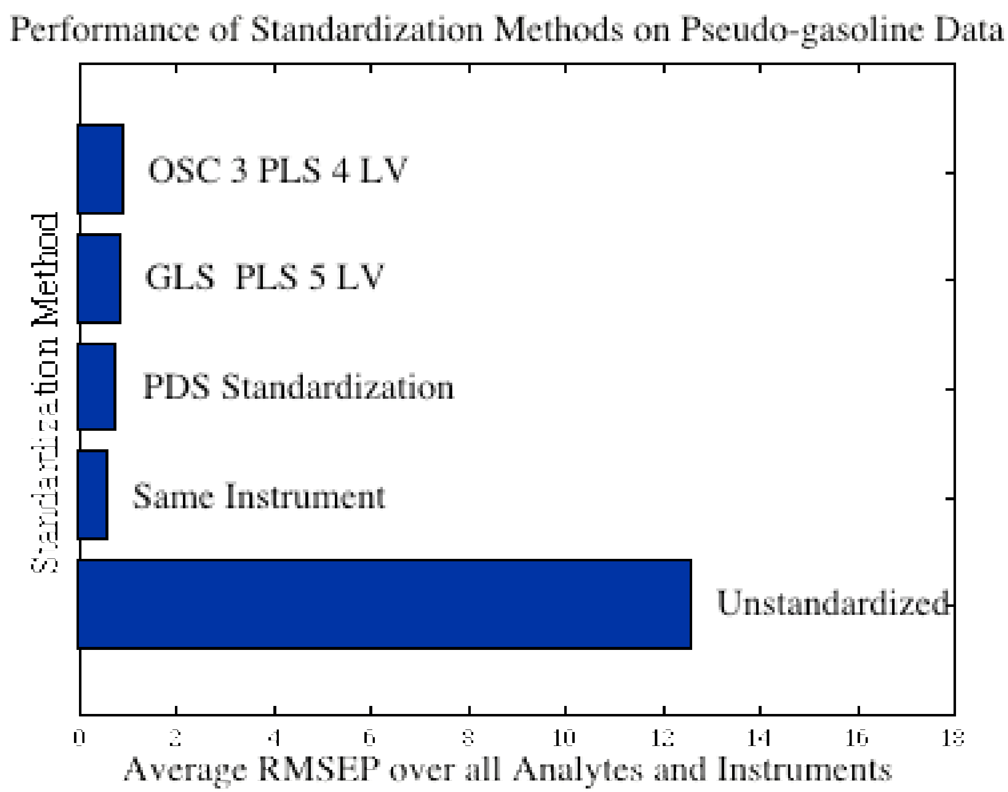
Results on Corn Data



Comparison of Methods on Pseudo Gasoline Data

- Available data
  - 30 samples split 20/10
  - 5 analytes
  - 2 instruments
- 5 Transfer samples selected
  - Based on model inverse for PDS
  - Based on PCA leverage for OSC, GLS
- Tested all 3 methods on all combinations of instrument and analyte

Results on Pseudo Gasoline Data



EIGENVECTOR RESEARCH INCORPORATED

Other Ways to Apply GLS

- GLS weighting may be applied directly to model
  - Don’t have to rebuild model!
  - Works well sometimes, but not always (future work)
- Downweight interferents
  - Requires estimate of effect of interferent
  - Image decluttering
- Upweight analyte of interest

EIGENVECTOR RESEARCH INCORPORATED

Usability Issues

	Meta-parameters?	Requires Y?	Rebuild calibration model?	Modifies spectra?	Transfer sets function of Y?	Affects net analyte signal?
GLS	1	No	Yes/No	Yes	No	Yes
PDS	2	No	No	No	Yes	No
OSC	3	Yes	Yes	Yes	No	Yes

Conclusions

- GLS preprocessing is a simple, effective method for eliminating spectral differences
- Can be used in several ways
- Only one adjustable parameter
- Potential loss of net analyte signal

Bibliography

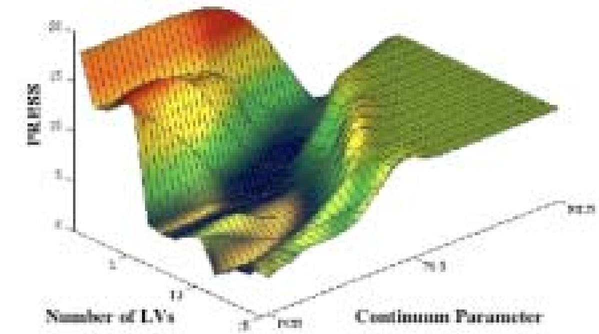
[1] H. Martens, M. Høy, B.M. Wise, R. Bro and P.B. Brockhoff, “GLS Preprocessing of Multivariate Data,” submitted to J. Chemometrics, May 2001.

[2] Y. Wang, D.J. Veltkamp and B.R. Kowalski, “Multivariate Instrument Standardization,” Anal. Chem., 63(23), pps 2750-2756, 1991.

[3] Z. Wang, T. Dean and B.R. Kowalski, “Additive Background Correction in Multivariate Instrument Standardization,” Anal. Chem., 67(14), pps 249-260, 1995.

[4] S. Wold, H. Antti, F. Lindgren and J. Öhman, “Orthogonal Signal Correction of Near-Infrared Spectra,” Chemo. and Intell. Lab. Sys., 44, pps 175-185, 1998.

[5] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg and S. Wold, “An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra,” Chemo. and Intell. Lab. Sys., 44, pps 229-244, 1998.



PLS_Toolbox 3.0

for use with MATLAB™

Barry M. Wise
Neal B. Gallagher
Rasmus Bro
Jeremy M. Shaver

EIGENVECTOR
RESEARCH INCORPORATED

Contact Information

Eigenvector Research, Inc.
830 Wapato Lake Road
Manson, WA 98831
Phone: (509)687-2022
Fax: (509)687-7033
Email: bmw@eigenvector.com
Web: eigenvector.com

EIGENVECTOR RESEARCH INCORPORATED