Combining Calibration Transfer and Preprocessing: What Methods, What Order?

Charles E. Miller, Robert T. Roginski,
Neal B. Gallagher and Barry M. Wise

Eigenvector Research, Inc.

Wenatchee, WA

©Copyright 2008
Eigenvector Research, Inc.
No part of this material may be photocopied or reproduced in any form without prior written consent from Eigenvector Research, Inc.

EIGENVECTOR
RESEARCH INCORPORATED

Outline

- The Calibration Transfer/ Instrument Standardization problem
- Why *this* study?
- Study Design:
  - Two Data sets (Cargill-Corn, Amoco pseudo-gas)
  - Three Standardization approaches (PDS, GLS and PDS-GLS)
  - Three Preprocessing approaches (2^{nd} d, MSC, SNV)
- Results Analysis (using PARAFAC!)
- Conclusions

EIGENVECTOR RESEARCH INCORPORATED

The Calibration Transfer Problem

- Some calibrations depend on very small instrument response changes
- No two instrument systems have identical responses!
  - Differences in hardware, optics, ***sampling interface***…
- **AND**…Temporal changes in a ***single*** instrument
  - Aging parts, mechanical hysteresis, dust/dirt buildup
  - Environmental conditions (T, humidity, vibration..)
- **AND**…New chemical/physical interferences
- Complete re-calibration or calibration updating is costly, not feasible

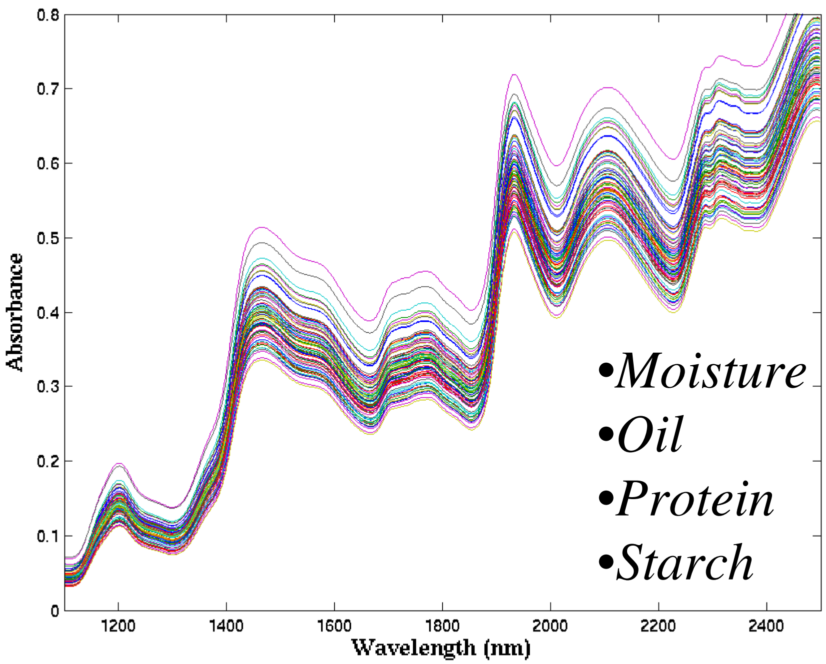
EIGENVECTOR RESEARCH INCORPORATED

Why this Study?

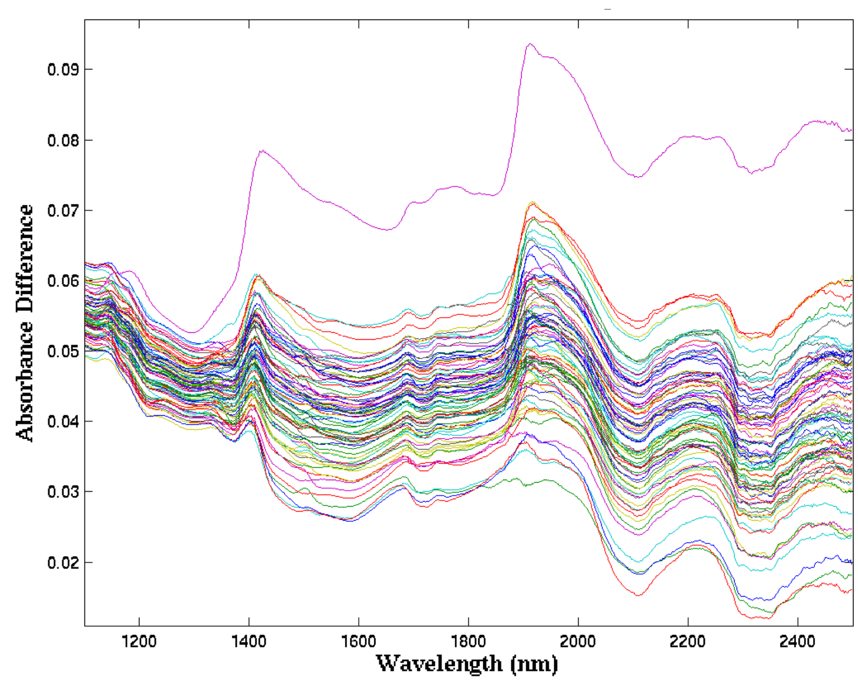
- Explore two industrially-relevant data structures
  - NIR transmission and diffuse reflectance
- Consider *hybrid* standardization strategies
- Explore *interactions* between common *preprocessing and standardization* methods
  - Could preprocessing be sufficient in some cases?

Corn Data

Instrument 1 data only



Sample-wise differences between instruments



80 samples analyzed on 3 different Foss NIRSystems instruments:

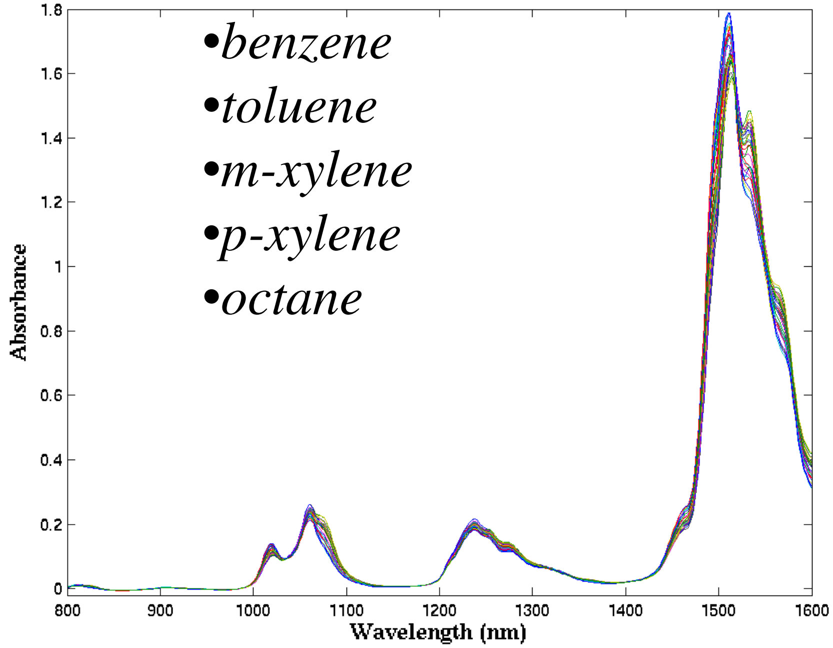
- **location A:** one model 6500 ("mp6"), and one model 5000 ("mp5")
- **location B:** another model 5000 ("m5")

From Mike Blackburn, Cargill

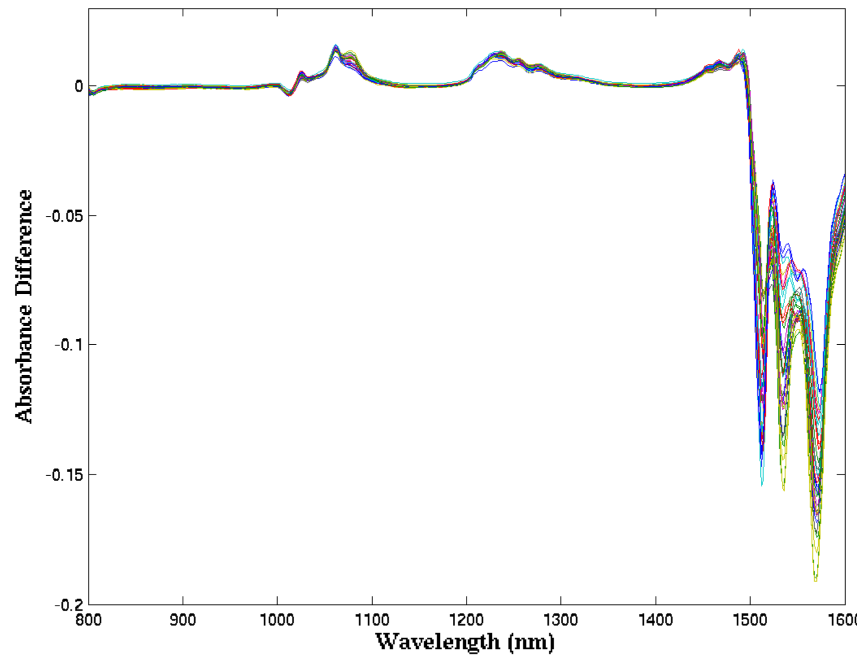
EIGENVECTOR RESEARCH INCORPORATED

Pseudo Gasoline Data

Instrument 1 data



Sample-wise differences between instruments

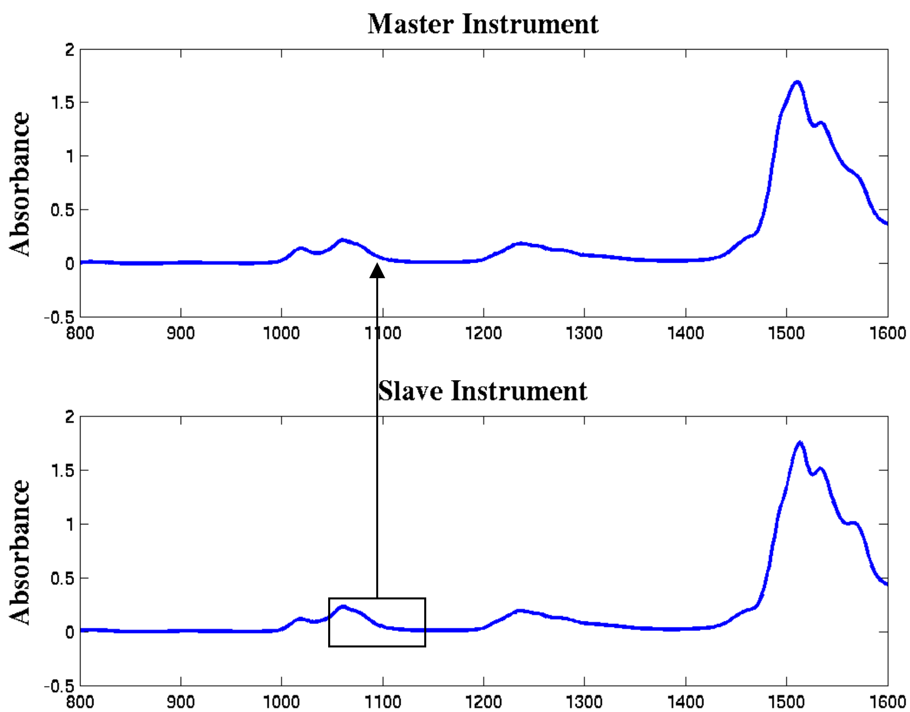


30 synthetic mixtures of 5 analytes analyzed on 2 different LT Industries instruments, with cuvette transmission interfaces

From Ernie Baughman, University of LaVerne (formerly at Amoco)

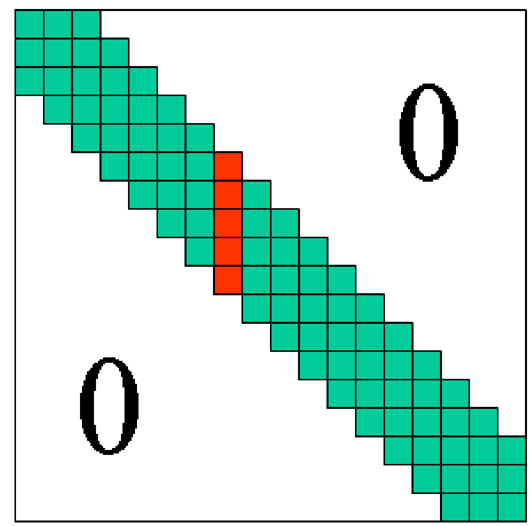
EIGENVECTOR RESEARCH INCORPORATED

Piece-wise Direct Standardization (PDS)



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

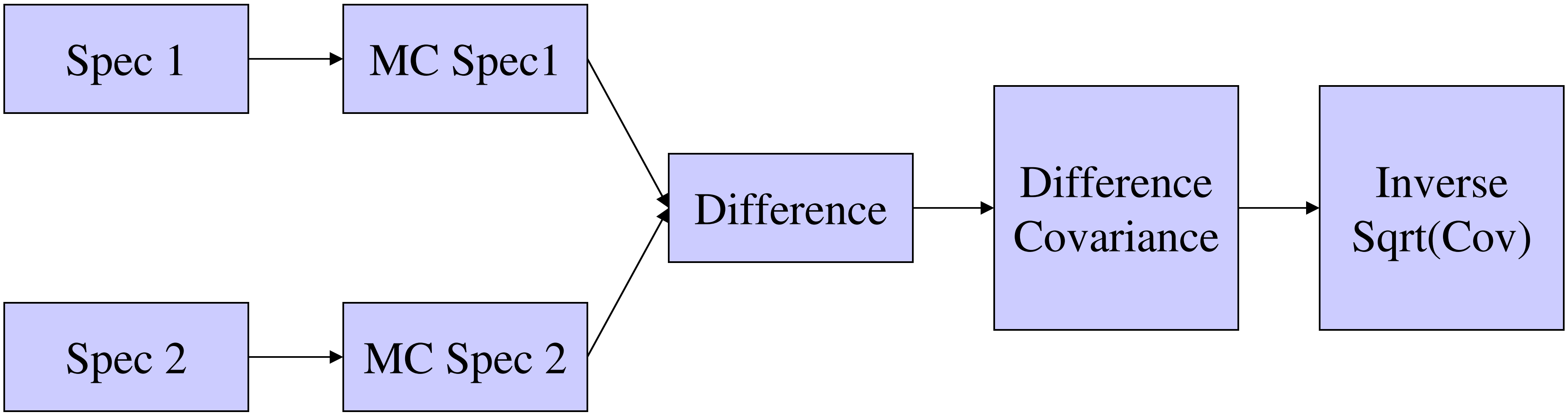
$$F =$$



*Several **adjacent** variables on Slave "mapped" to a **single** variable on Master (handles wavelength registration shifts!)*

EIGENVECTOR RESEARCH INCORPORATED

GLS Weighting



"down-weight" variables that reveal differences between the instruments (reduce variation that the instruments do NOT have in common!)

EIGENVECTOR RESEARCH INCORPORATED

Experimental Details

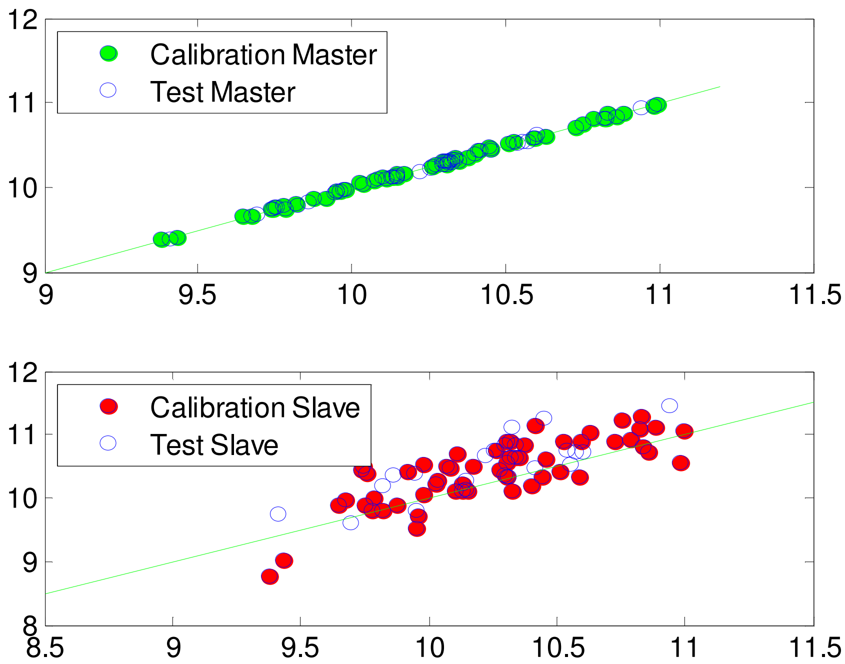
- Calibration/Test splits: 20/10 for pseudo-gas, 60/20 for Corn
  - *Manually* selected to span analyte concentration space
- GLS:
  - Down-weighting parameter = 1e-6
  - 10 transfer standards, based on *distance from mean*
- PDS:
  - Window width = 5 pts, or 10nm
  - 10 transfer standards, based on *leverage of model built using master calibration data*
- Model Building: SIMPLS on *slave calibration* set
  - Cross-validation: venetian blinds, 8 subsets
  - Optimal number of LVs determined from RMSECV curve (by CEM)
- Model Testing: RMSEP of model applied to test set
- Preprocessing: 2nd der., MSC (mean), SNV
  - 2nd der.: order = 2, width = 11 pts (22 nm)

EIGENVECTOR RESEARCH INCORPORATED

Sample results- corn data

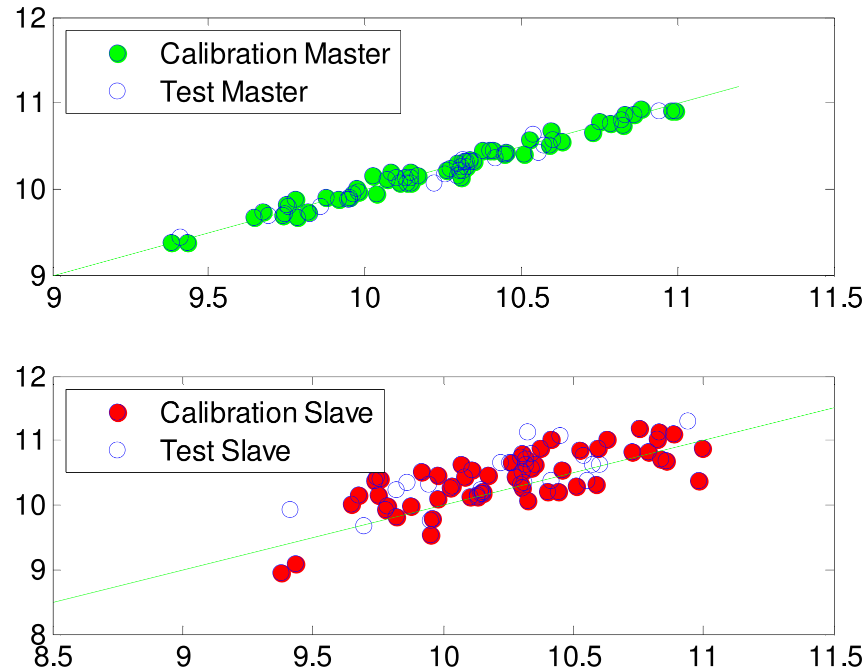
Analyte = Moisture, m5 master/mp5 slave

PDS only



8LVs, RMSEP = 0.415

PDS, then 2nd derivative



7LVs, RMSEP = 0.392

(2 of 672 cases!...)

EIGENVECTOR RESEARCH INCORPORATED

Results Table (only for Corn, PDS)

Corn PREDICTION RESULTS-PDS				No preprocessing (PDS only)		MSC AFTER PDS		SNV AFTER PDS		2nd deriv AFTER PDS		MSC BEFORE PDS		SNV BEFORE PDS		2nd deriv BEFORE PDS	
Analyte	master	slave	index	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV	RMSEP Test Slave	#LV
1	m5	mp5	112	0.416	9	0.246	6	0.246	6	0.386	9	0.264	6	0.245	6	0.411	7
	m5	mp6	113	0.480	9	0.271	6	0.271	6	0.464	9	0.260	6	0.260	6	0.430	7
	mp5	m5	211	0.380	10	0.226	6	0.249	8	0.309	8	0.487	8	0.237	6	0.319	8
	mp5	mp6	213	0.219	10	0.209	6	0.237	8	0.158	8	0.240	8	0.214	6	0.161	8
	mp6	m5	311	0.349	9	0.304	5	0.304	5	0.224	8	0.344	5	0.269	5	0.197	8
	mp6	mp5	312	0.109	9	0.260	5	0.260	5	0.115	8	0.267	5	0.260	5	0.131	8
2	m5	mp5	122	0.085	8	0.135	4	0.107	5	0.098	9	0.106	4	0.107	4	0.082	9
	m5	mp6	123	0.069	8	0.128	4	0.128	4	0.083	9	0.100	4	0.099	4	0.080	9
	mp5	m5	221	0.092	7	0.079	5	0.079	5	0.080	5	0.078	5	0.083	5	0.055	5
	mp5	mp6	223	0.095	7	0.087	5	0.087	5	0.072	5	0.082	5	0.082	5	0.065	5
	mp6	m5	321	0.082	8	0.074	5	0.074	5	0.100	6	0.077	5	0.079	5	0.061	6
	mp6	mp5	322	0.096	7	0.087	5	0.087	5	0.077	6	0.089	5	0.090	5	0.065	6
3	m5	mp5	132	0.188	9	0.239	8	0.227	8	0.213	7	0.195	8	0.183	8	0.205	7
	m5	mp6	133	0.197	9	0.235	8	0.234	8	0.181	7	0.192	8	0.197	8	0.181	7
	mp5	m5	231	0.170	8	0.182	6	0.206	8	0.141	5	0.266	8	0.202	6	0.155	5
	mp5	mp6	233	0.219	8	0.176	8	0.168	8	0.157	5	0.232	8	0.158	8	0.179	5
	mp6	m5	331	0.166	8	0.205	6	0.205	6	0.157	5	0.362	6	0.208	6	0.167	5
	mp6	mp5	332	0.204	8	0.175	6	0.175	6	0.190	5	0.175	6	0.181	6	0.194	5
4	m5	mp5	142	0.350	9	0.403	7	0.400	6	0.448	7	0.503	6	0.406	8	0.456	5
	m5	mp6	143	0.371	9	0.413	6	0.415	6	0.449	8	0.402	6	0.366	8	0.413	7
	mp5	m5	241	0.316	8	0.303	8	0.263	8	0.299	5	0.469	8	0.313	8	0.365	5
	mp5	mp6	243	0.417	8	0.328	8	0.332	8	0.347	5	0.321	8	0.320	8	0.378	5
	mp6	m5	341	0.395	8	0.179	7	0.183	7	0.369	5	0.201	7	0.303	7	0.424	5
	mp6	mp5	342	0.458	9	0.367	7	0.360	7	0.416	5	0.366	7	0.369	7	0.422	5
averages analyte 1				0.325	9.3	0.253	5.7	0.261	6.3	0.276	8.3	0.310	6.3	0.247	5.7	0.275	7.7
averages analyte 2				0.086	7.5	0.098	4.7	0.093	4.8	0.085	6.7	0.089	4.7	0.090	4.7	0.068	6.7
averages analyte 3				0.191	8.3	0.202	7.0	0.202	7.3	0.173	5.7	0.237	7.3	0.188	7.0	0.180	5.7
averages analyte 4				0.384	8.5	0.332	7.2	0.326	7.0	0.388	5.8	0.377	7.0	0.346	7.7	0.410	5.3
averages over analytes				0.247	8.4	0.221	6.1	0.221	6.4	0.231	6.6	0.253	6.3	0.218	6.3	0.233	6.3

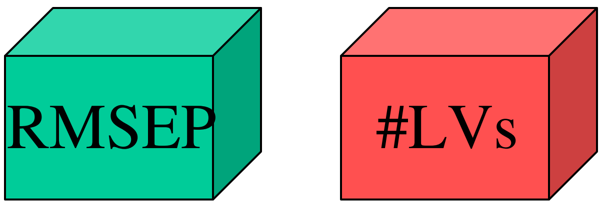
This is ONE way to explore results……

EIGENVECTOR RESEARCH INCORPORATED

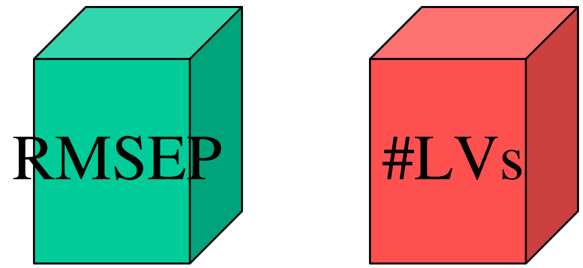
Result Arrays- two for each dataset!

- *RMSEP of prediction on* ***test set***
- *Number of LVs needed in the* ***calibration***

Corn - 672 cases (4x7x6x4)



Pseudo-Gas: 280 cases (4x7x2x5)



- Mode 1: standardization (4 values)
  - None, PDS, GLS, PDS followed by GLS
- Mode 2: preprocessing (7 values)
  - None, and 2D, MSC or SNV- *before and after* standardization
- Mode 3: master/slave pairing (2 for gas, 6 for corn)
- Mode 4: analyte (5 for gas, 4 for corn)

EIGENVECTOR RESEARCH INCORPORATED

PARAFAC on results

- A 1-component PARAFAC model explores *4^{th} order interactions* between *all 4 effects*:
  - Standardization method
  - Preprocessing method
  - Master/slave pairing, and
  - Analyte
- GEMANOVA : generalized multiplicative ANOVA – includes lower-order interactions
  - *R. Bro, M. Jakobsen, J. Chemometrics* 2002; 16 294-304

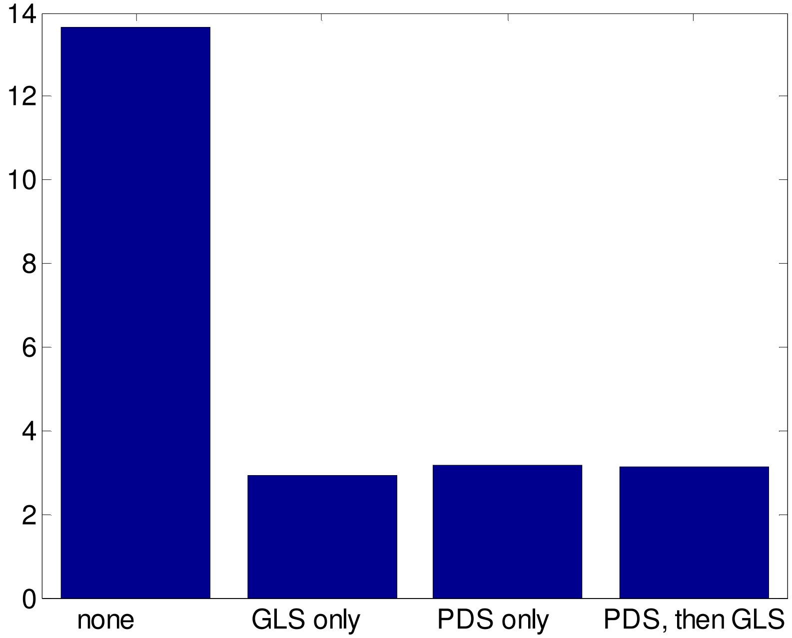
Explained variance: 1-component PARAFAC model

	Corn data	Pseudo gas data
RMSEP results	91.3	82.2
optimal #LV results	96.9	96.8

EIGENVECTOR RESEARCH INCORPORATED

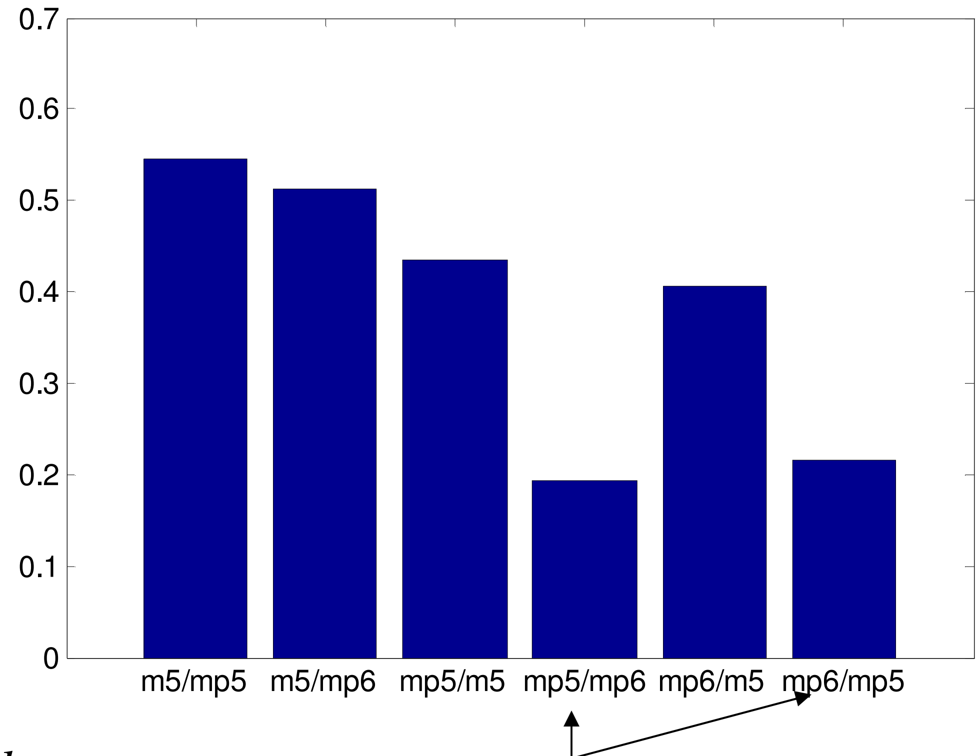
PARAFAC loadings- RMSEP, Corn

Mode 1 (standardization method)



Some standardization ***definitely*** *needed for Corn data!; GLS slightly out-performs PDS and PDS-GLS*

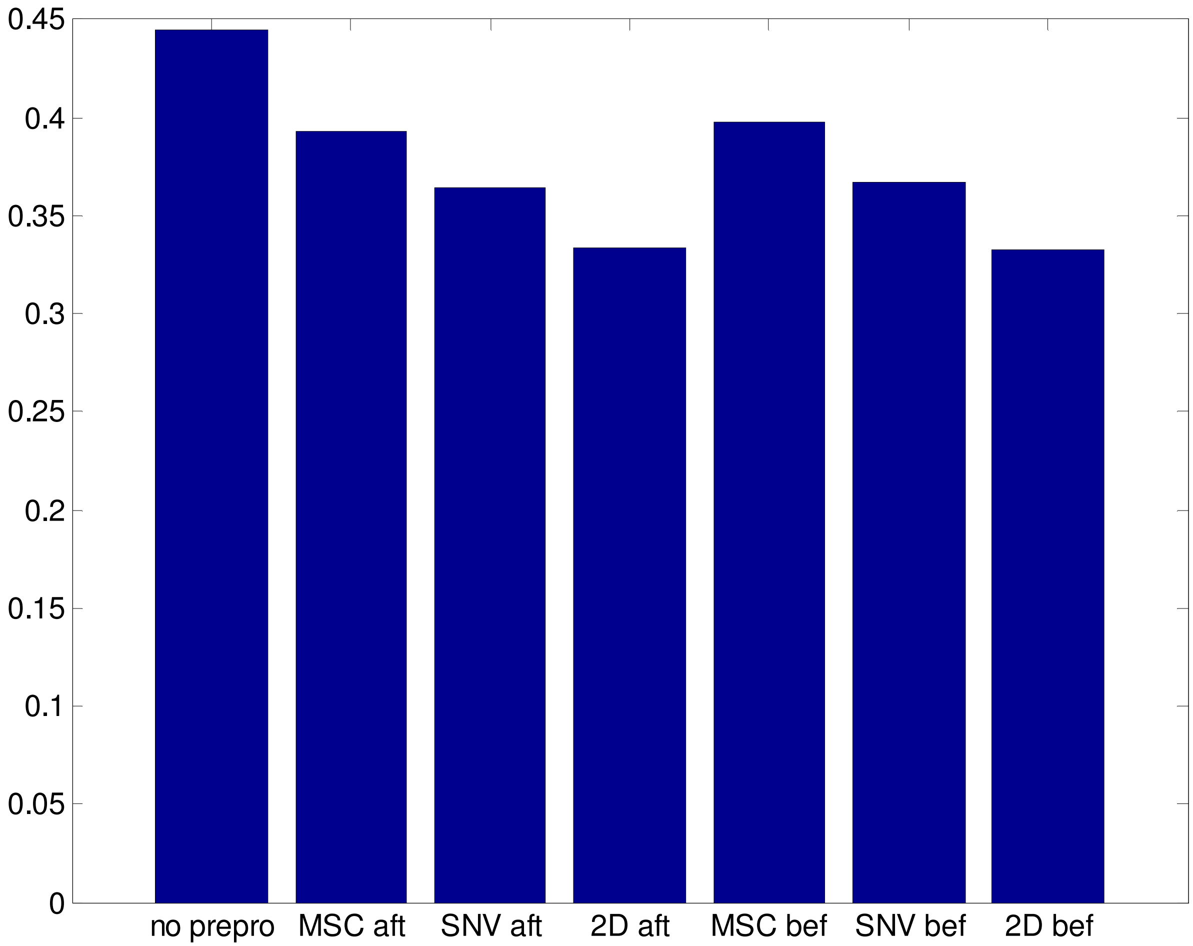
Mode 3 (master/slave instrument pair)



All transfers ***within the same location*** *result in lower RMSEPs!*

EIGENVECTOR RESEARCH INCORPORATED

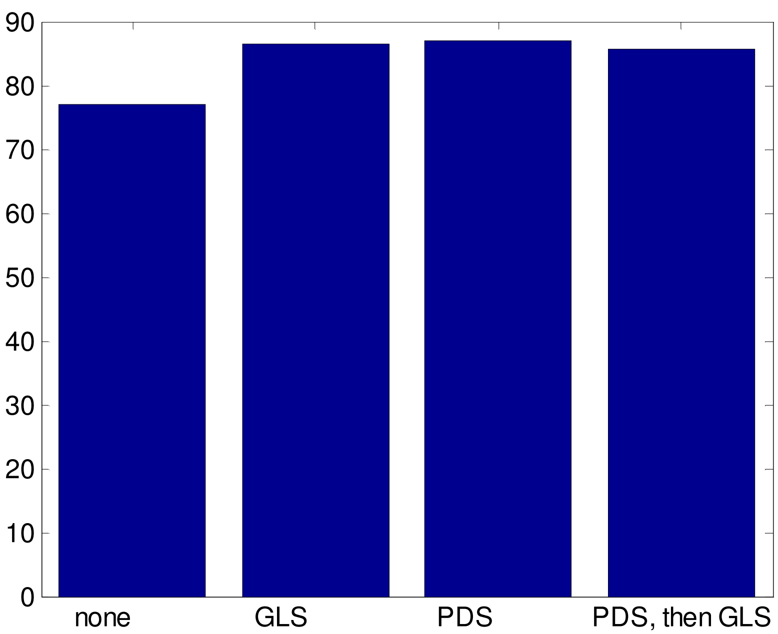
PARAFAC loadings- RMSEP, Corn

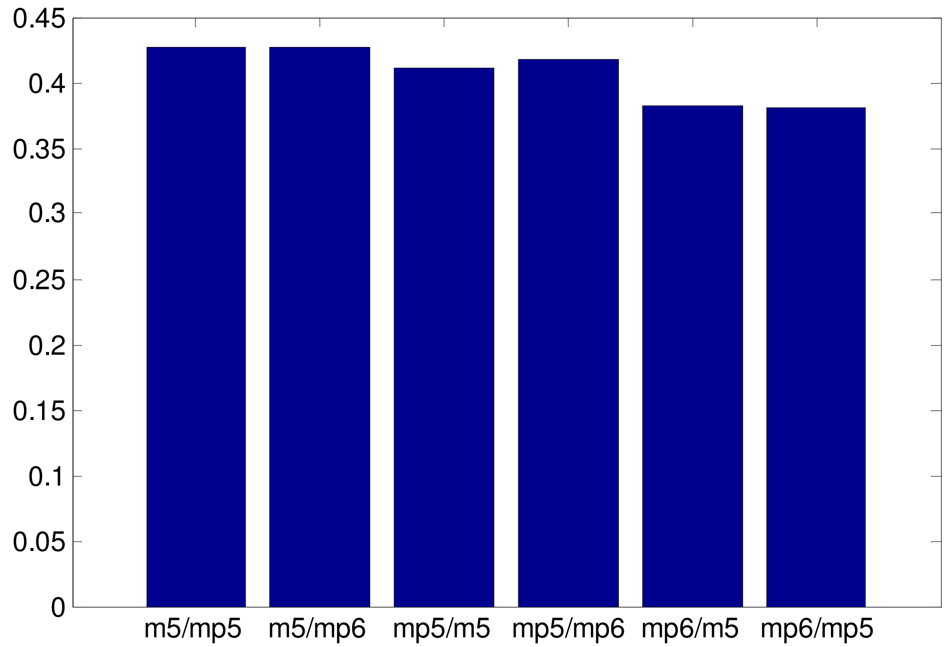


- *All preprocessing types reduce RMSEP*
- *2nd derivative generates lowest RMSEP*
- *preprocessing **after** standardization SLIGHTLY better*

EIGENVECTOR RESEARCH INCORPORATED

PARAFAC loadings- #LVs, Corn





*Slightly **more** LVs required for standardized data vs. non-standardized data !*

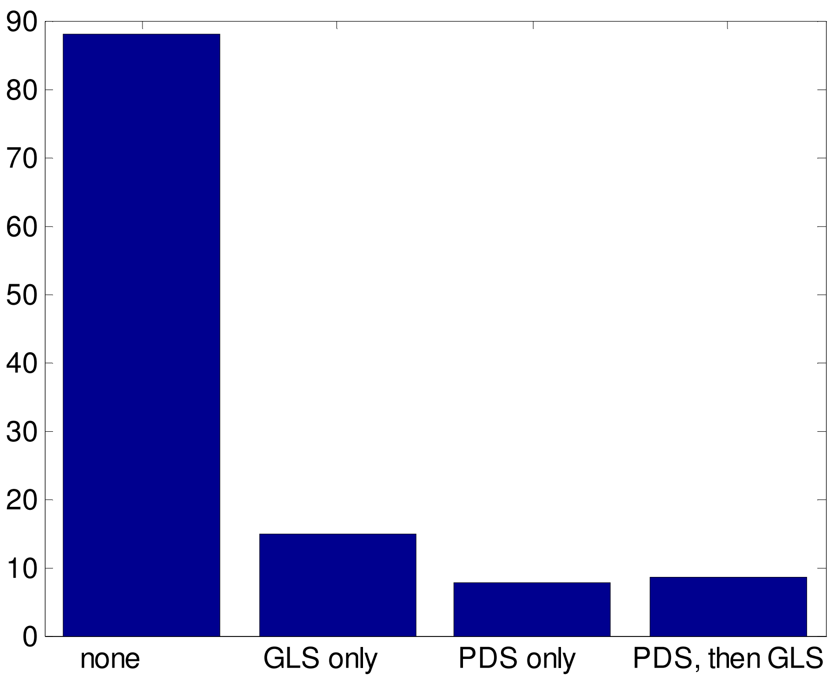
*Transfers in which **the 6500 instrument (mp6) is the master** require fewer LVs!*

Interesting!...

EIGENVECTOR RESEARCH INCORPORATED

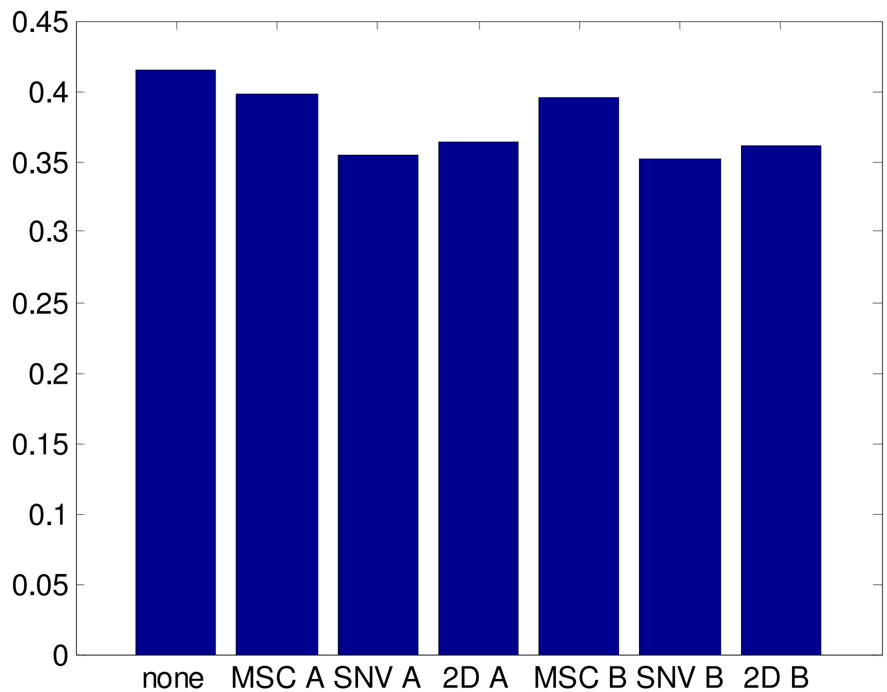
PARAFAC loadings- RMSEP, Gas

Mode 1 (standardization method)



Some standardization definitely needed for Gas data!; PDS and hybrid method out-perform GLS

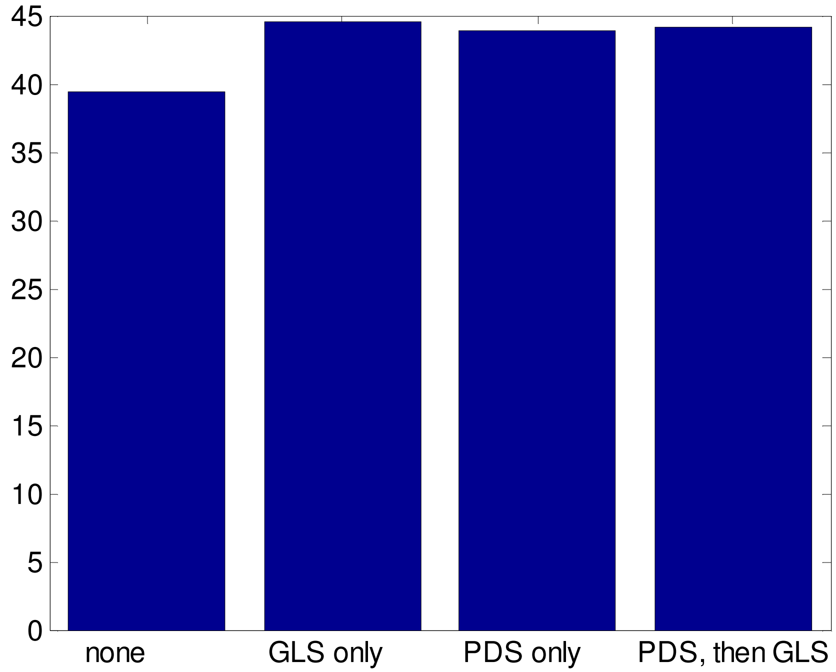
Mode 2 (preprocessing)



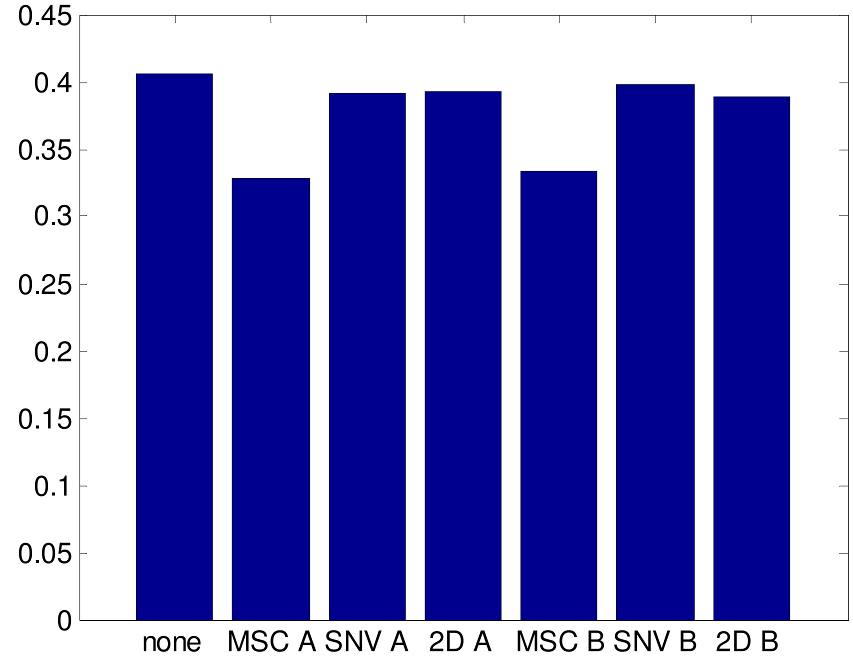
*SNV generates lowest RMSEP; preprocessing **before** standardization SLIGHTLY better*

EIGENVECTOR RESEARCH INCORPORATED

PARAFAC loadings- #LVs, Gas



*Once again- slightly **more** LVs required for standardized data vs. non-standardized data!*



Only MSC (before or after standardization) reduces the required number of LVs in the model

EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- Standardization is needed in both cases
- Hybrid PDS-GLS method **not** much different than PDS alone
- Preprocessing before vs. after standardization makes little difference
  - 2nd D best for Corn, SNV best for Gas
- Corn data (diffuse reflectance)
  - GLS slightly out-performs PDS, PDS-GLS
  - Transfers within the same location give lowest RMSEPs
    - *Even though they're different instrument **models**!*
- Pseudo-gas data (transmission)
  - PDS, PDS-GLS clearly out-perform GLS
- Slightly **more** LVs required for standardized data vs. non-standardized data

Future Work

- Expand study to include
  - Orthogonal Signal Correction (OSC), FIR and other standardization methods
  - Other “hybrids” of standardization methods
  - Data from other instrument types, sampling interfaces
- GLS options
  - Weighting could be applied *directly to model*
  - Down-weight interferences, up-weight analytes

EIGENVECTOR RESEARCH INCORPORATED

Contact Information

Eigenvector Research, Inc.
3905 West Eaglerock Drive
Wenatchee, WA 98801
Phone: (509)662-9213
Fax: (509)662-9214
Email: bmw@eigenvector.com
Web: eigenvector.com

EIGENVECTOR RESEARCH INCORPORATED