Integrating Instrument Standardization Methods into Data Preprocessing Schemes

Barry M. Wise and R. Scott Koch

Eigenvector Research Inc., Wenatchee, WA, USA

EIGENVECTOR RESEARCH INCORPORATED

Abstract

Multivariate calibration, classification and fault detection are ubiquitous in the monitoring and control of chemical and pharmaceutical processes. Model maintenance can be defined as the on-going servicing of these multivariate models in order to preserve their predictive capabilities. It is required because of changes to either the sample matrices or the instrument response. The goal of model maintenance is to sustain or improve models over time and changing conditions with the least amount of cost and effort. Instrument standardization methods are an important element in the model maintenance toolbox. The goal of instrument standardization is to map the response of a separate instrument, or the same instrument at a later time, onto the response of the master/standard instrument. Many methods have been developed for instrument standardization, including Direction Standardization (DS), many variations on Piecewise Direct Standardization (PDS), and Spectral Subspace Transformation (SST). Multivariate calibration models often include a number of preprocessing steps before the actual regression, classification or other model is applied. But how should the standardization method be integrated with the preprocessing scheme? Should standardization be done before or after preprocessing? Or even between preprocessing steps? Our experience suggests that this question does not have a universal answer, and the optimal approach is case-specific. Based on this we have developed a framework for standardization that allows insertion into calibration models before, after, or in between preprocessing steps. This preprocessing/standardization framework is presented in this talk, and several representative cases are demonstrated.

EIGENVECTOR RESEARCH INCORPORATED

Instrument Standardization is an Aspect of Model Maintenance

- Numerous things can cause calibration models to become invalid
  - samples move to a range outside original calibration
  - new variation is introduced into the samples
  - a change in the sample matrix causes the relationship between analyte and measurement to change
    - change in pH, particle size, temperature
  - a change in the hardware causes the analyte-measurement relationship to change
    - entirely different instrument, maintenance, fiber optic change, source replacement, etc.
- Last two items are often handled with instrument standarization (aka calibration transfer)

EIGENVECTOR RESEARCH INCORPORATED

Before Model Goes Online

- Develop a plan for maintenance
  - Assume that updated or new calibration models will eventually be required
  - Have a plan for how to detect the problem and what to do about it
  - Put it in the budget!
- Measure standard samples
  - Plan for registration and amplitude shifts
  - Characterize instrument in ranges important to model

EIGENVECTOR RESEARCH INCORPORATED

Detecting Model/Data Mismatch & Performance Degradation

- Model prediction diagnostics
  - Spectral residual Q (or similar)
  - Sample distance T^2 (or similar)
- Prediction accuracy monitored via primary reference method
  - Unlikely that change not detected by diagnostics but possible
  - Risk based approach?
- Detecting that *something* has gone wrong easier than determining *what* has gone wrong.

EIGENVECTOR RESEARCH INCORPORATED

Standardization Methods

- Many methods available to estimate the response of the standard instrument from a different or changed instrument
- My favorites
  - Direct Standardization (DS)
  - Piecewise Direct Standardization (PDS)
    - and double window PDS (DWPDS)
  - Spectral Subspace Transform (SST)
  - Generalized Least Squares Preprocessing (GLS)
  - Orthogonal Signal Correction (OSC)

EIGENVECTOR RESEARCH INCORPORATED

Standardization Methods

Method	Number of meta-parameters	Y values *not* required?	Use original calibration model?	Spectra un-modified?	Transfer sets *not* function of Y?	Retains net analyte signal?	Can use generic standards?	Number transfer samples required
DS	1							High
PDS	2				✗			Low
SST	1				✗			Medium
GLS	1		✗	✗		✗		Medium
OSC	2-3	✗	✗	✗	✗	✗	✗	Medium

Use original calibration model with transformed slave data

EIGENVECTOR RESEARCH INCORPORATED

Data Preprocessing

- Preprocessing is done to reduce extraneous variance so that relevant variance can be more easily modeled
- Data preprocessing is a part of most calibration and classification models
- Preprocessing often consists of multiple steps
  - E.g. MSC, followed by 1st derivative, followed by mean centering

Where to place Standardization Transform?

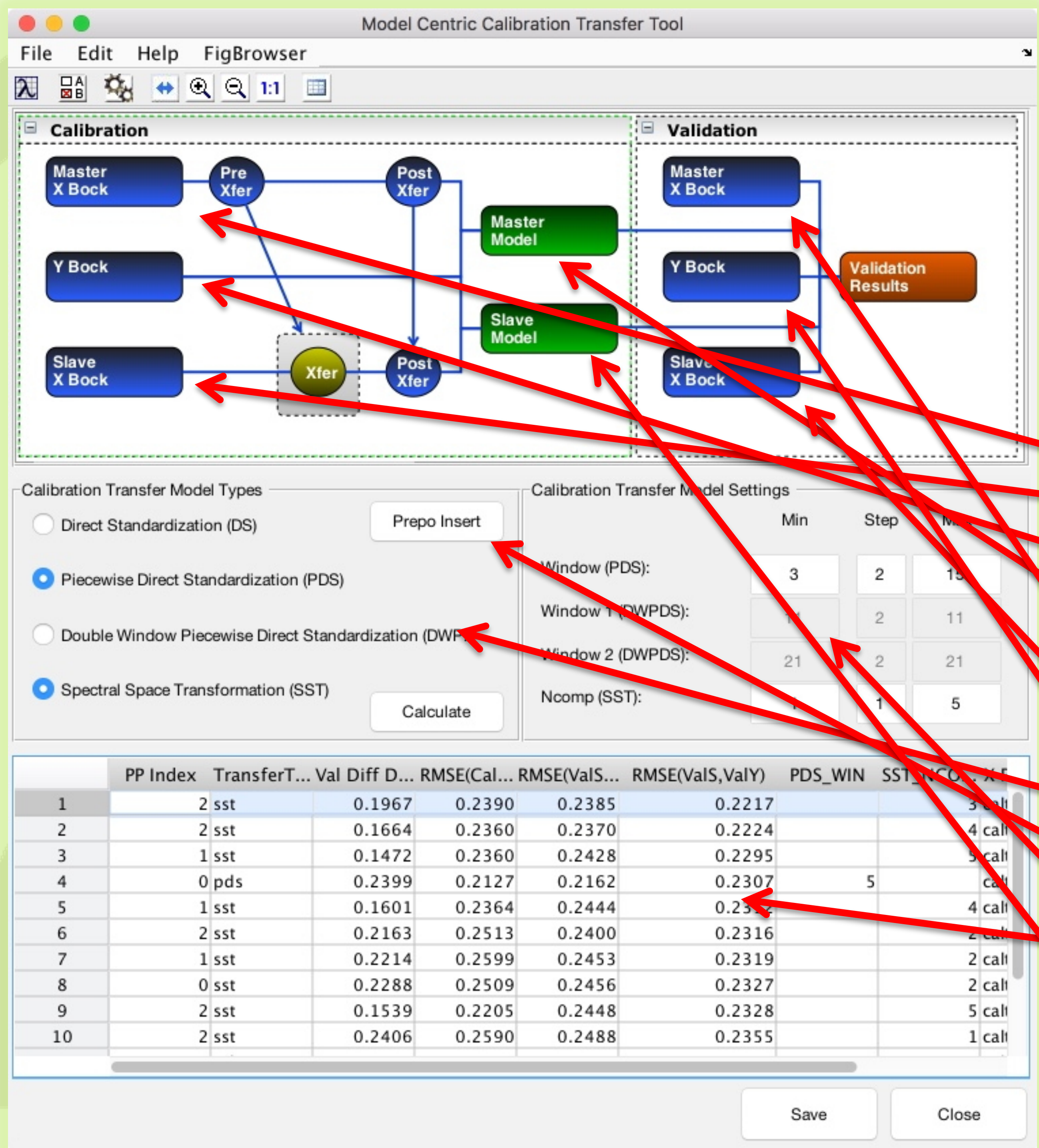
- Standardization transforms are most often done on raw data
- But why not after preprocessing or even between steps?
- May be advantages to standardization after preprocessing
  - Preprocessing designed to reduce unwanted variability
  - Preprocessing my also help in the identification and application of standardization transforms

EIGENVECTOR RESEARCH INCORPORATED

Model Centric Calibration Transfer (MCCT)

- Calculate standardization transforms between steps in existing calibration models
- Evaluate performance of transfer on calibration and validation sets on a variety of metrics (spectral difference, y-prediction error, etc.)
- Inserts standardization into existing model for use on slave instruments

EIGENVECTOR RESEARCH INCORPORATED



MCCT Interface

Data/Model Elements:

- Master transfer X
- Slave transfer X
- Calibration common Y
- Master model
- Master validation X
- Slave validation X
- Validation common Y

Standardization Methods

Insertion Points

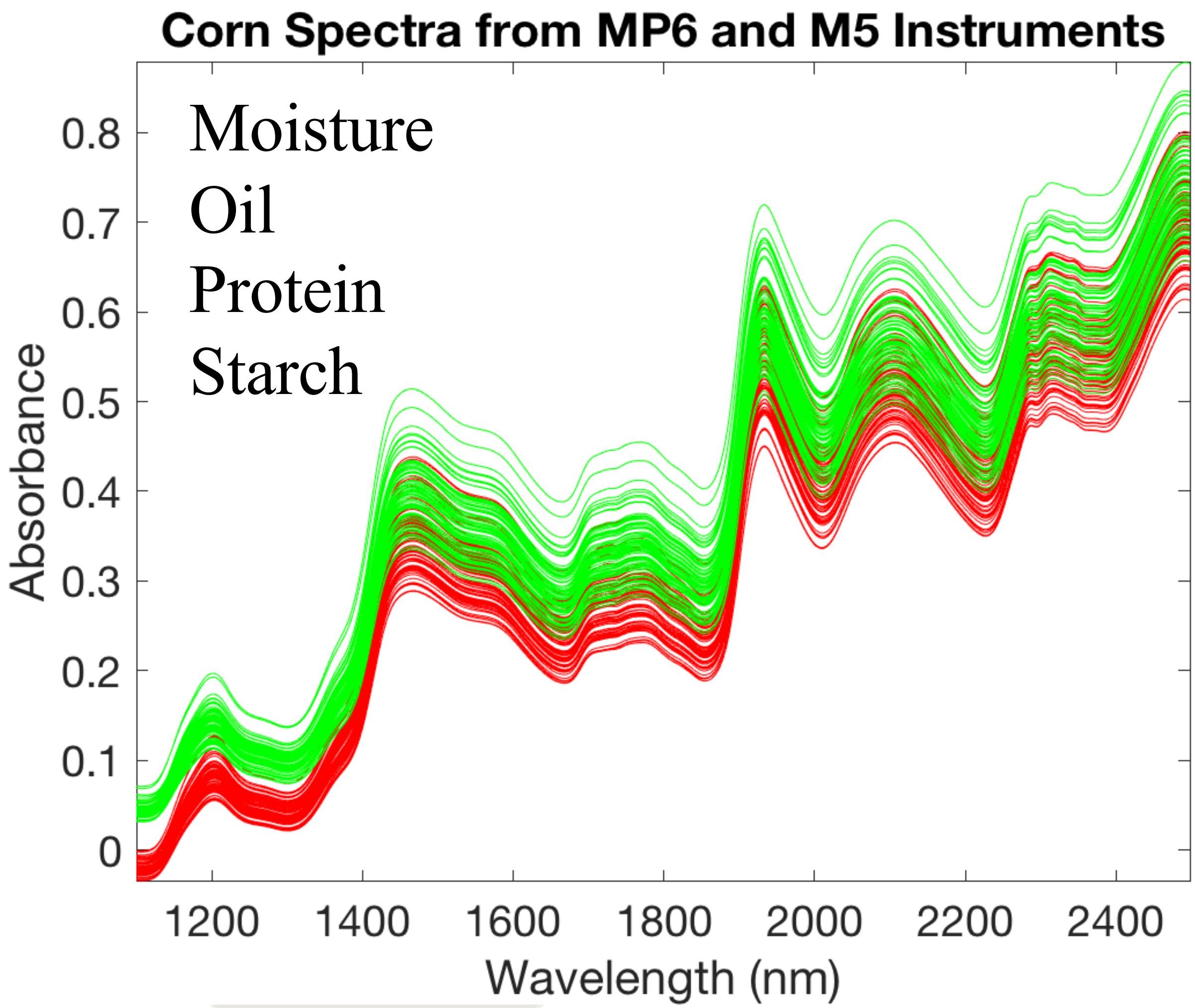
Method Parameters

Validation Results

Calculated Slave Model

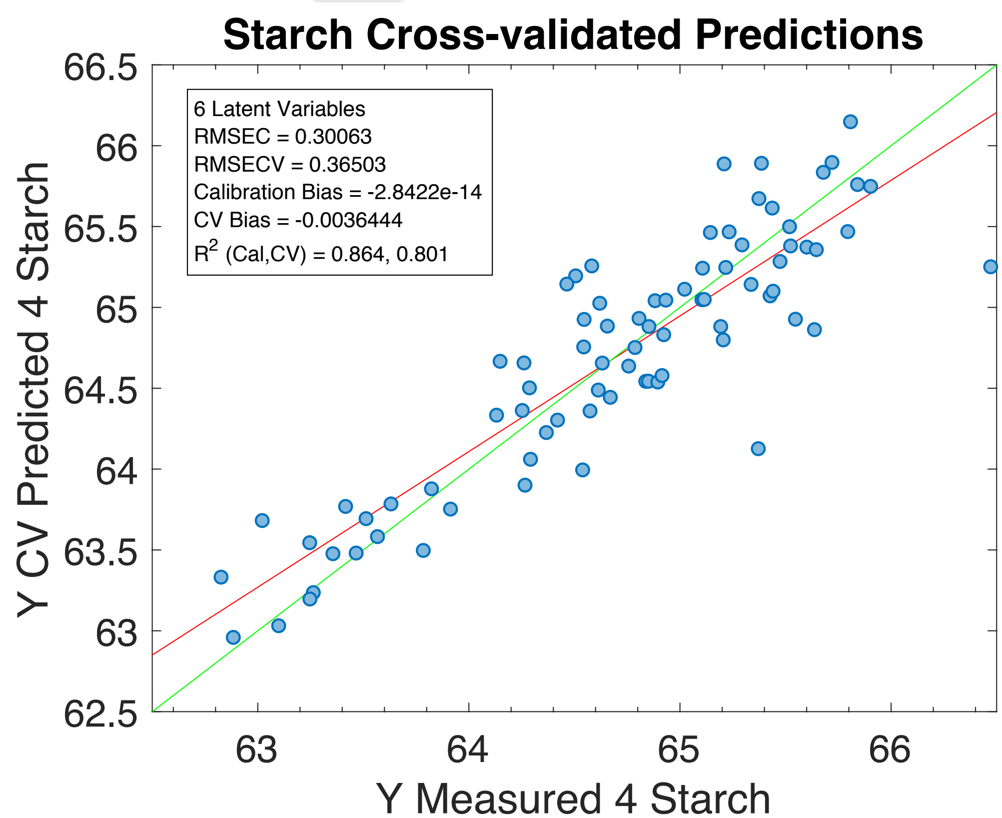
EIGENVECTOR RESEARCH INCORPORATED

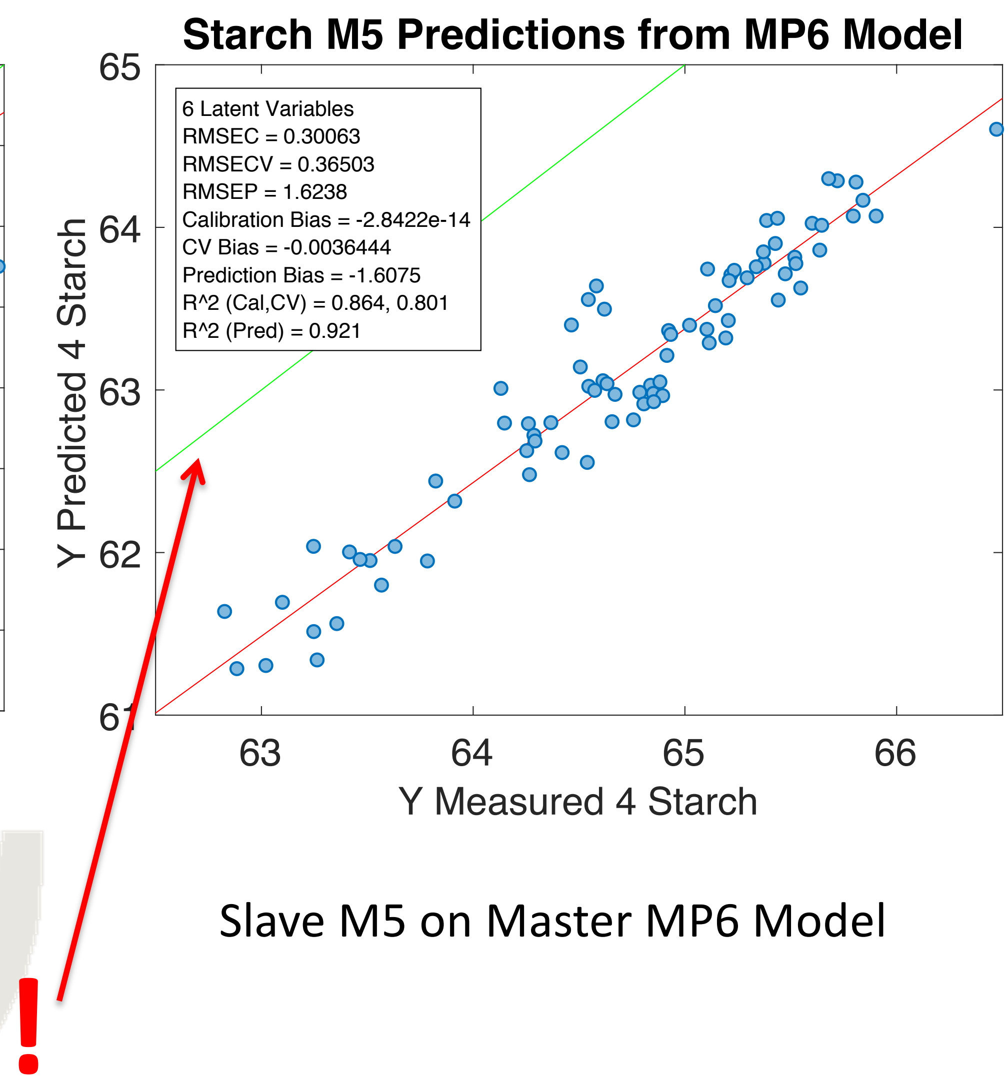
Example: Corn Data



Data provided by Cargill

Starch Model Predictions





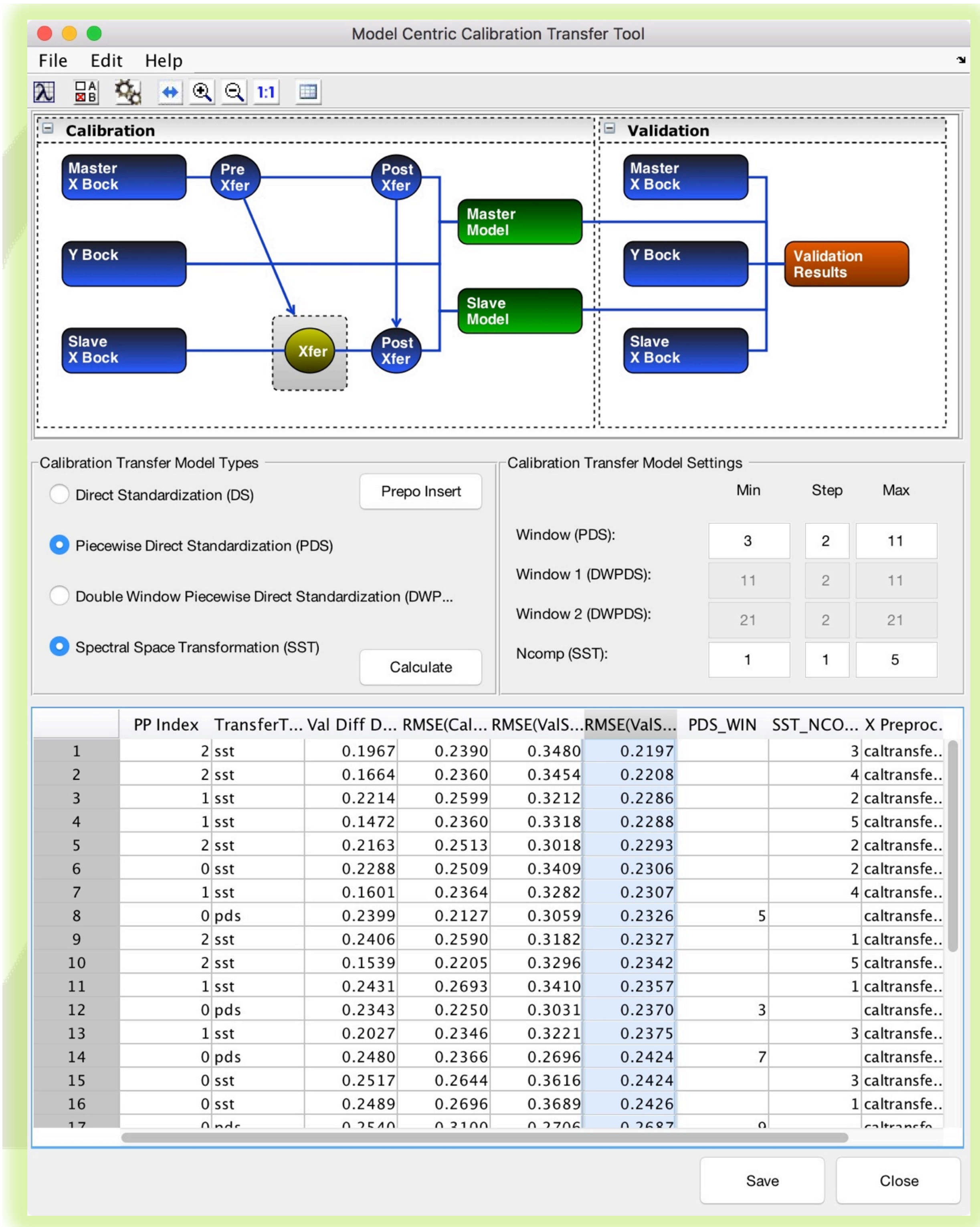
Master MP6 Calibration Preprocessing

- MSC
- 1st Derivative
- Mean centering

Slave M5 on Master MP6 Model

EIGENVECTOR RESEARCH INCORPORATED

MCCT Results for Corn Starch

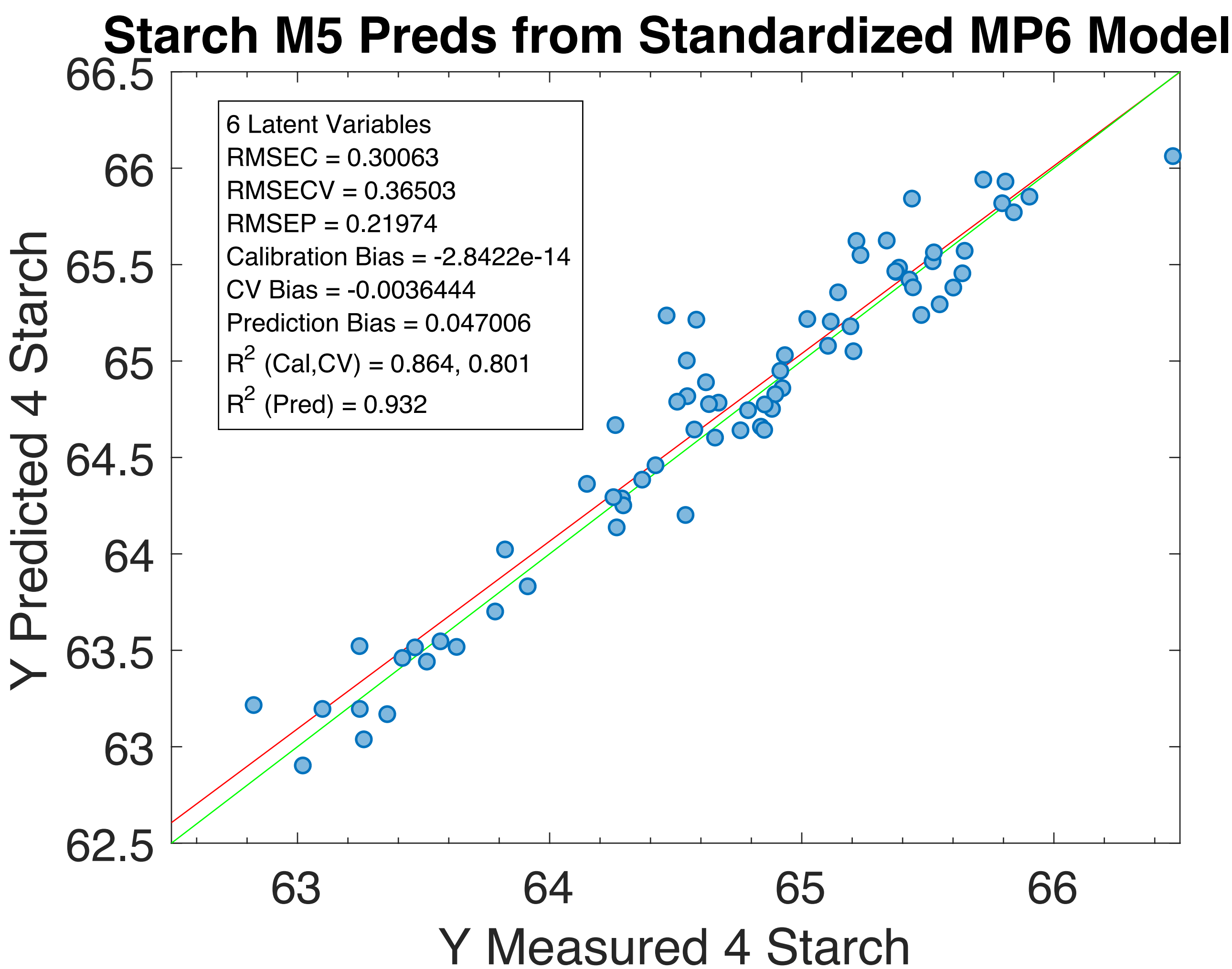


	PP Index	TransferT...	Val Diff D...	RMSE(Cal...	RMSE(ValS...	RMSE(ValS...	PDS_WIN	SST_NCO...	X Preproc.
1	2	sst	0.1967	0.2390	0.3480	0.2197		3	caltransfe..
2	2	sst	0.1664	0.2360	0.3454	0.2208		4	caltransfe..
3	1	sst	0.2214	0.2599	0.3212	0.2286		2	caltransfe..
4	1	sst	0.1472	0.2360	0.3318	0.2288		5	caltransfe..
5	2	sst	0.2163	0.2513	0.3018	0.2293		2	caltransfe..
6	0	sst	0.2288	0.2509	0.3409	0.2306		2	caltransfe..
7	1	sst	0.1601	0.2364	0.3282	0.2307		4	caltransfe..
8	0	pds	0.2399	0.2127	0.3059	0.2326	5		caltransfe..
9	2	sst	0.2406	0.2590	0.3182	0.2327		1	caltransfe..
10	2	sst	0.1539	0.2205	0.3296	0.2342		5	caltransfe..
11	1	sst	0.2431	0.2693	0.3410	0.2357		1	caltransfe..
12	0	pds	0.2343	0.2250	0.3031	0.2370	3		caltransfe..
13	1	sst	0.2027	0.2346	0.3221	0.2375		3	caltransfe..
14	0	pds	0.2480	0.2366	0.2696	0.2424	7		caltransfe..
15	0	sst	0.2517	0.2644	0.3616	0.2424		3	caltransfe..
16	0	sst	0.2489	0.2696	0.3689	0.2426		1	caltransfe..
17	0	pds	0.2540	0.3100	0.2706	0.2687	9		caltransfe..

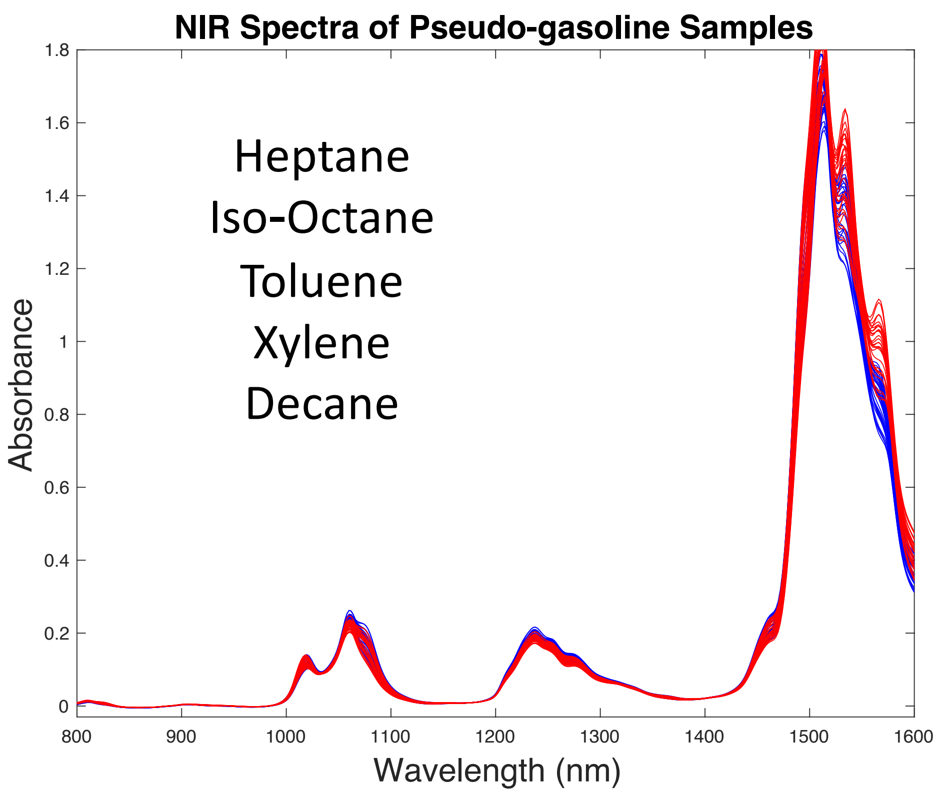
- Table ordered by prediction error on validation set
- Best transfers are SST inserted between prepro steps

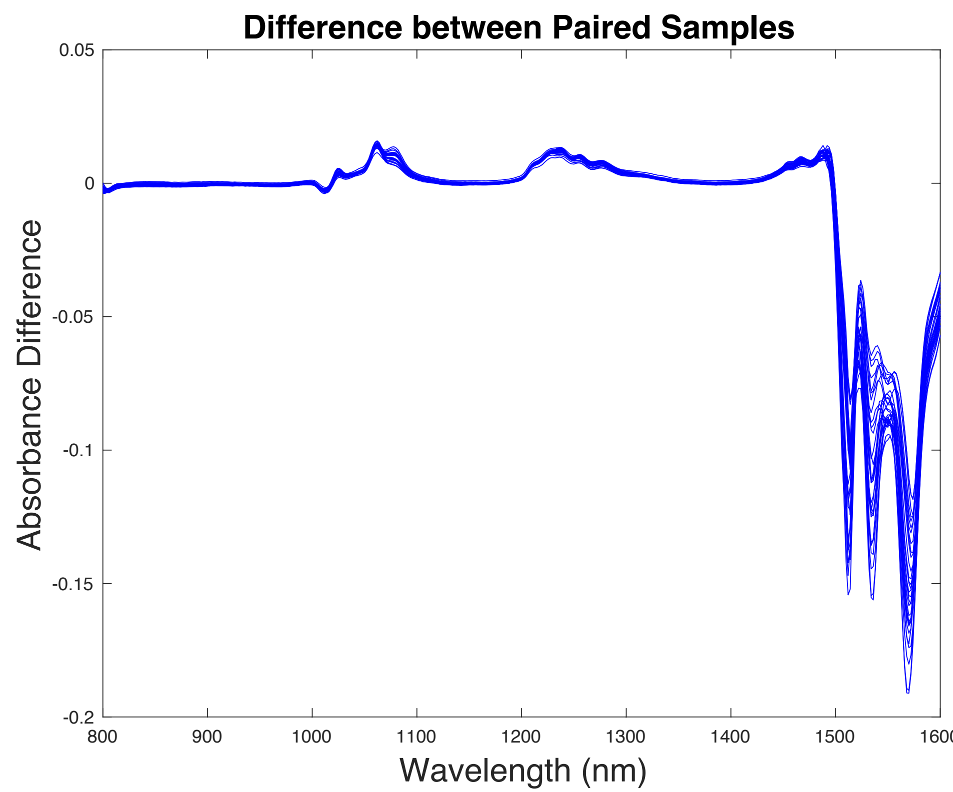
EIGENVECTOR RESEARCH INCORPORATED

Predictions from Standardized Model



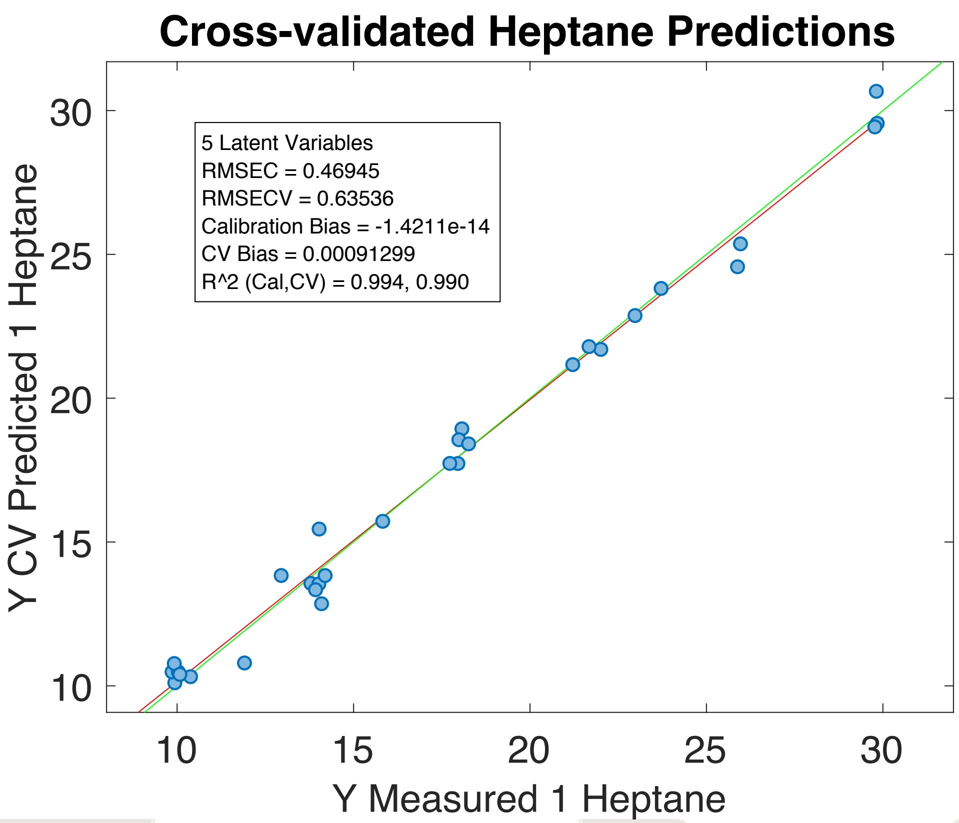
Example: NIR of Pseudo-gasolines

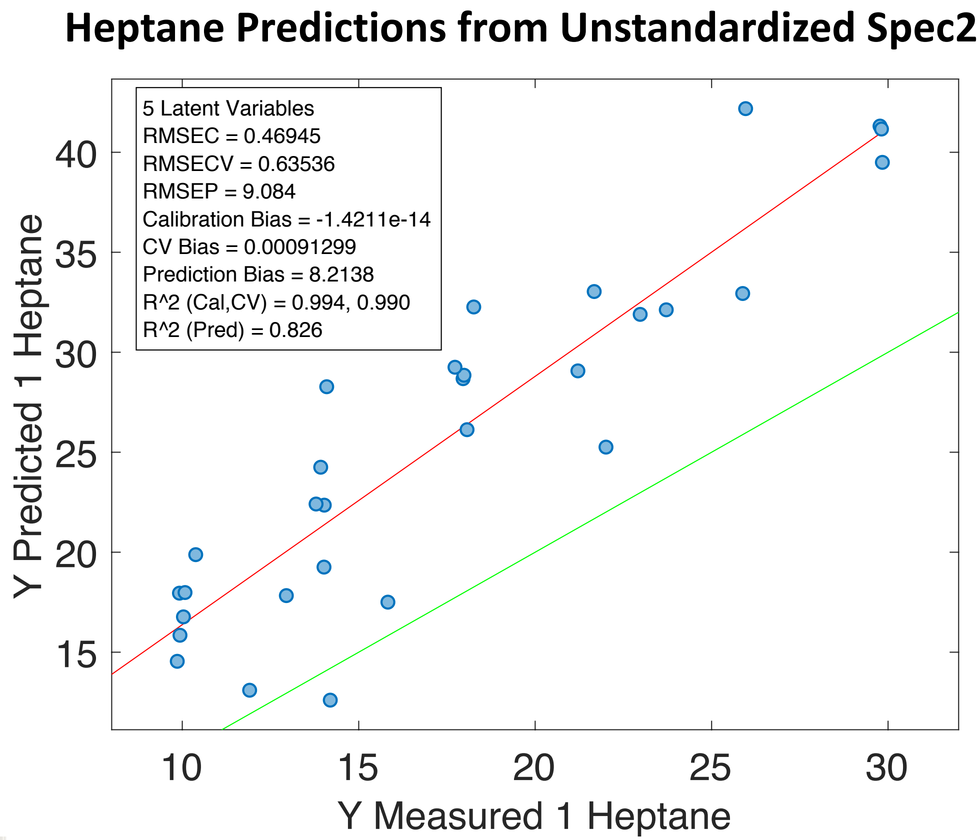




Data provided by AMOCO

Heptane Calibration Model Predicitions





Master Spec1 Calibration

Preprocessing

- Mean centering

Slave Spec2 on Master Spec1 Model

EIGENVECTOR RESEARCH INCORPORATED

Testing Standardization Methods

Calibration Transfer Model Types

- Direct Standardization (DS)
- Piecewise Direct Standardization (PDS)
- Double Window Piecewise Direct Standardization (DWPDS)
- Spectral Space Transformation (SST)

Prepo Insert

Calculate

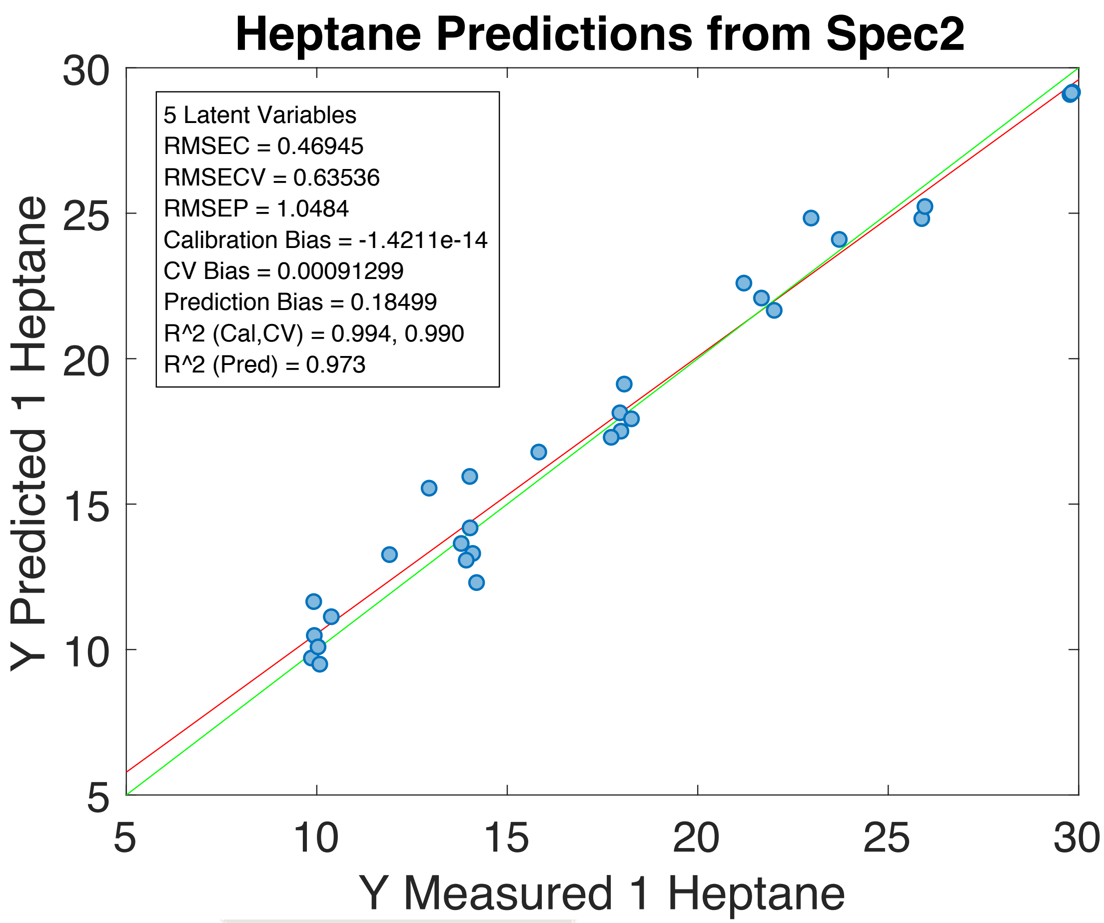
Calibration Transfer Model Settings

	Min	Step	Max
Window (PDS):	1	2	11
Window 1 (DWPDS):	1	2	5
Window 2 (DWPDS):	5	8	29
Ncomp (SST):	1	1	5

	PP ...	Trans...	Val Diff Data	RMSE(CalS,CalY)	RMSE(ValS,ValM)	RMSE(ValS,ValY)	PDS_WIN	DWPDS_W...	DWPDS_W...	SST_NCO...	X Prepro
1	0	pds	0.0308	0.7266	1.0460	1.1016	3				caltransf
2	0	pds	0.0347	0.7750	1.2039	1.2562	1				caltransf
3	0	dwpds	0.0356	0.6462	1.2105	1.2623		1	5		caltransf
4	0	pds	0.0271	1.0345	1.2920	1.2724	11				caltransf
5	0	dwpds	0.0331	0.9337	1.4011	1.4192		5	5		caltransf
6	0	pds	0.0300	0.7517	1.4426	1.4429	5				caltransf
7	0	dwpds	0.0444	1.0342	1.5898	1.6424		1	13		caltransf
8	0	dwpds	0.0336	0.9356	1.7264	1.7166		3	5		caltransf
9	0	pds	0.0278	1.2534	1.7456	1.7500	9				caltransf
10	0	dwpds	0.0412	1.1703	1.7544	1.7704		5	21		caltransf
11	0	dwpds	0.0506	1.5400	1.7014	1.7880		1	21		caltransf
12	0	dwpds	0.0386	1.2121	1.7834	1.7957		5	13		caltransf
13	0	pds	0.0287	1.2813	1.8972	1.8931	7				caltransf
14	0	dwpds	0.0417	1.4759	2.0065	2.0314		3	21		caltransf
15	0	dwpds	0.0402	1.3491	2.1076	2.1174		3	13		caltransf
16	0	sst	6.6223e-15	0.4945	2.6061	2.5836				5	caltransf
17	0	dwpds	0.0433	1.5469	2.5253	2.5965		3	29		caltransf

EIGENVECTOR RESEARCH INCORPORATED

Heptane Predictions from Standardized Spec2



EIGENVECTOR RESEARCH INCORPORATED

Using Standardized Models

- Create models with **PLS_Toolbox** or **Solo 8.2**
- Standardized models can be used just like normal models to make predictions from slave instruments
- **Solo_Predictor** stand-alone prediction engine
- Coming soon: standardized models exported as compile-able recipes with **Model_Exporter**

EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- Standardization an important aspect of model maintenance
- In some instances standardization transforms are best done after or between preprocessing steps
- MCCT is a platform for constructing/testing transforms and insertion points
- Creates easily implementable standardized models
- See: **http://tinyurl.com/MCCT-Tool**

EIGENVECTOR RESEARCH INCORPORATED