Model Maintenance—The Unrecognized Cost in PAT and QbD

Barry M. Wise and Robert T. Roginski

Eigenvector Research Inc., Wenatchee, WA, USA

Abstract

- Multivariate calibration, classification and fault detection models are ubiquitous in PAT and QbD. They occur in the both the development of processes and their permissible operating limits, (*i.e.* models for relating the process design space to product quality), and in manufacturing (*i.e.* models used in monitoring and control). Model maintenance can be defined as the on-going servicing of these multivariate models in order to preserve their predictive abilities. 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. This talk presents a roadmap (see figure below) for determining when model maintenance is required, the probable source of the response variations, and the appropriate approaches for achieving it. Methods for evaluating model robustness in order to identify models with lower ongoing maintenance costs are also discussed.

Outline

- Introduction to Model Maintenance
  - Definition and Goal
  - Causes of data/model mismatch
- Identifying the need for action
  - External validation samples
  - Model prediction diagnostics
    - Q and T^2
  - Setting action limits
  - Determining the source of the problem

Outline

- Model updating methods
  - Adding to the calibration set
  - Slope and bias adjustments
  - Automatic model updating
- Instrument standardization methods
  - Direct and Piecewise Direct standardization
  - Spectral Subspace Transformation
  - Filtering approaches: GLS and OSC

EIGENVECTOR RESEARCH INCORPORATED

Outline

- Avoiding model updating
  - "Brittleness" arising from preprocessing methods
  - Model robustness tests
- Conclusions

Definition and Goal

- Multivariate calibration, classification and fault detection models are ubiquitous in PAT and QbD.
- Model maintenance: The on-going servicing of multivariate models to preserve their predictive abilities.
- Goal of model maintenance: Sustain (or improve) models over time and changing conditions with the least amount of cost and effort

EIGENVECTOR RESEARCH INCORPORATED

Why Model Maintenance?

- Numerous things can cause multivariate models to become invalid
  - samples move to a range outside original calibration
    - analyte or interferent goes beyond calibration range or occurs in unusual combination
  - new variation is introduced into the samples
    - new interferent or variation in physical parameter, *e.g.* temperature
  - a change in the sample matrix causes the relationship between analyte and measurement to change
    - change in pressure, pH, particle size, temperature
  - a change in the hardware causes the analyte-measurement relationship to change
    - instrument maintenance, fiber optic change, source replacement, etc.

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.
  - unless you are monitoring via reference method!
  - ... and why aren't you?

EIGENVECTOR RESEARCH INCORPORATED

Diagnostics Limits

- By default, limits for Q and T^2 are generally provided based on some type of confidence limit
  - These limit values are statistically based
- Meaningful limits require knowledge of the process or measurement
  - Observations are needed that represent
    - "bad" states
    - out of spec product
    - failing sensor(s) or analyzer*

*these may be challenging to obtain

Setting Limits

- Often helpful to think in terms of failure modes
  - For processes, what are the undesirable states?
    - Keep in mind – we generally can't perform experiments at manufacturing scale
    - Is data at a smaller scale available?
      - Is this data conformable with manufacturing scale data?
  - Use these undesirable states as rational guides to setting limits on Q and T^2
    - Contributions are often helpful in identifying what type of fault
  - Sometimes modeling can provide additional insight

Example – Modeling to aid in Setting Limits

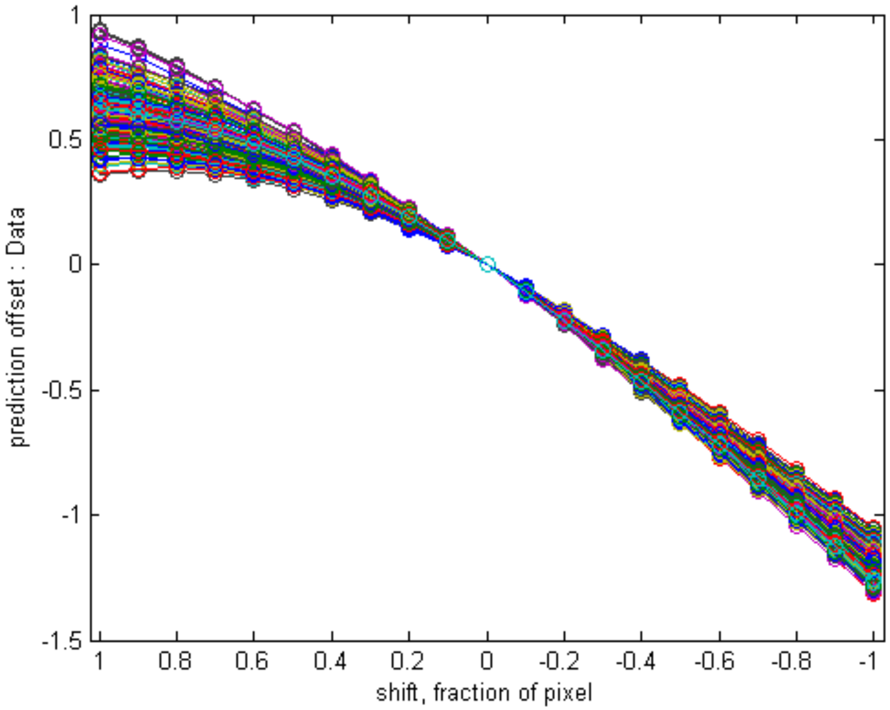
- Raman spectroscopic measurement of a solution to predict level of solute
  - Lab measurement
    - Controlled environment
- Failure modes
  - Sample
    - Solvent only
    - Empty cuvette
    - Wrong product
  - Very high values for Q and T^2 (> 8)
  - What about the measurement?

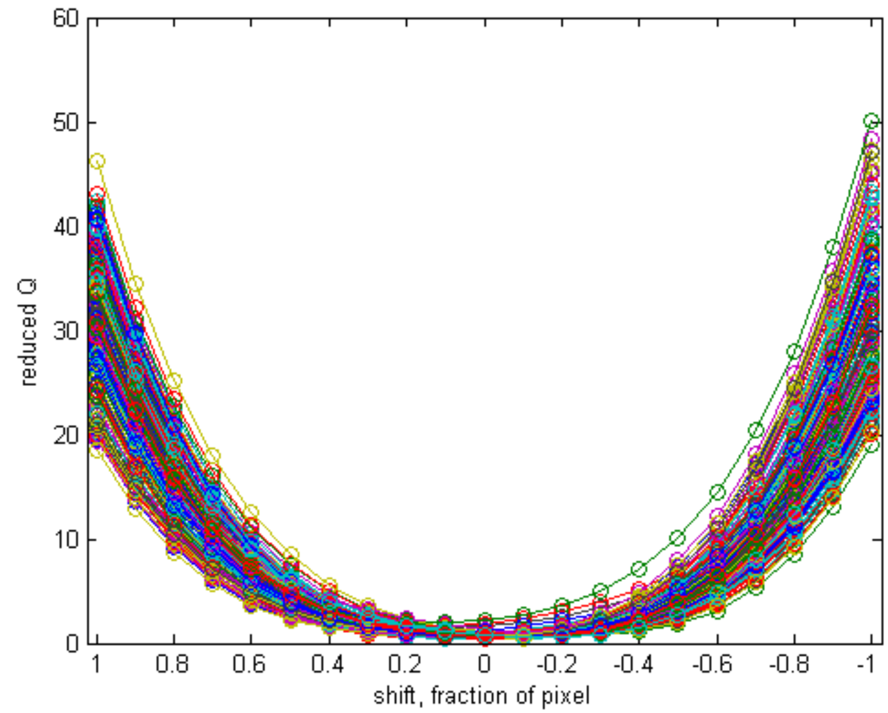
EIGENVECTOR RESEARCH INCORPORATED

Instrumental Issues

- Potential failure modes*
  - Significantly reduced power from laser
    - Validation samples showed no ill effect
  - Anecdotal evidence showed the system to be prone to x-axis registration instability
  - How does this impact the outputs of the model?
  - Approach
    - Use an existing validation set and generate pixel shifts ranging from -1 to 1 pixel in steps of 0.1 pixels
    - Apply model to each iteration
    - Determine impact on prediction, Q, and T^2

Impact

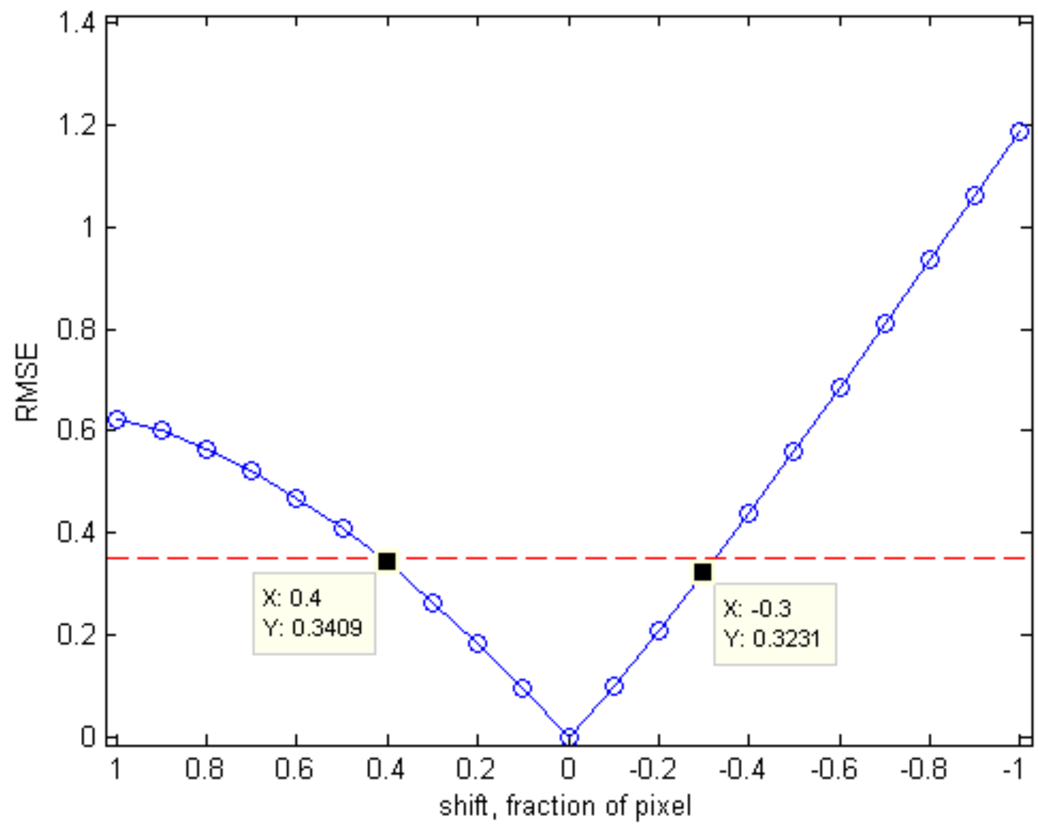

prediction offset : Data
shift, fraction of pixel


reduced Q
shift, fraction of pixel

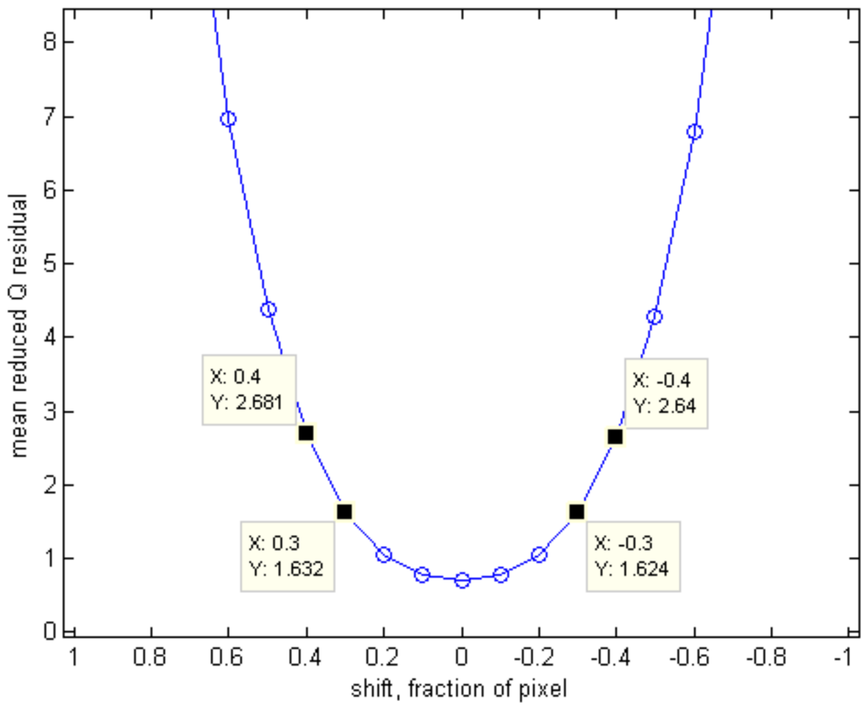
How to Select Limits

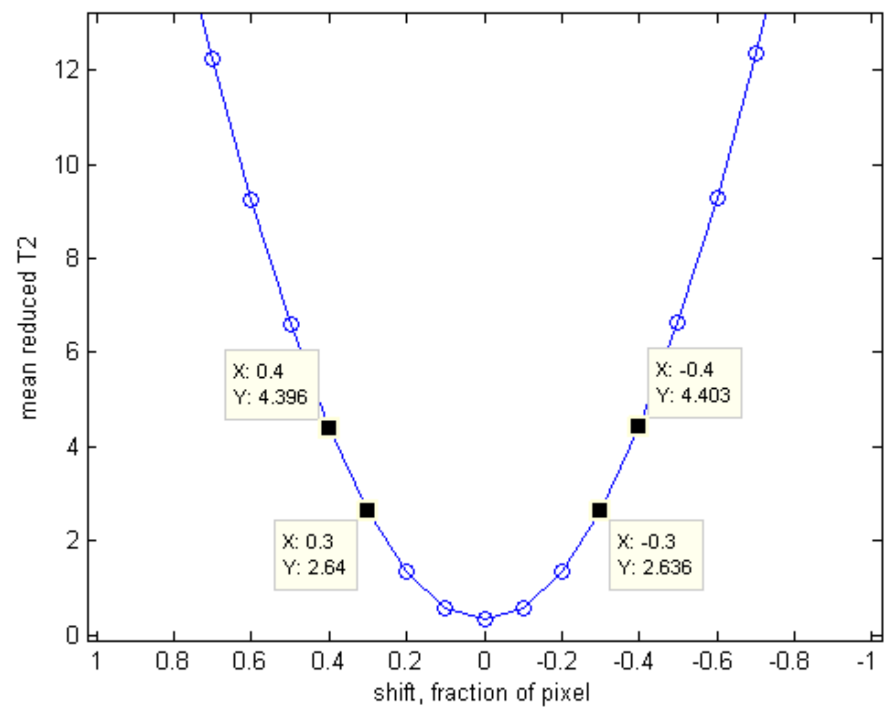
- Main consideration: impact on prediction error
  - Decision: use extreme values or summaries
    - Use RMSE for prediction error and mean values for Q and T^2
  - Choose an acceptable level of impact on prediction error
    - Corresponding values for Q and T^2

Impact on RMSE



Impact on Q and T^2





Q and T^2 impact curves fairly symmetric

EIGENVECTOR RESEARCH INCORPORATED

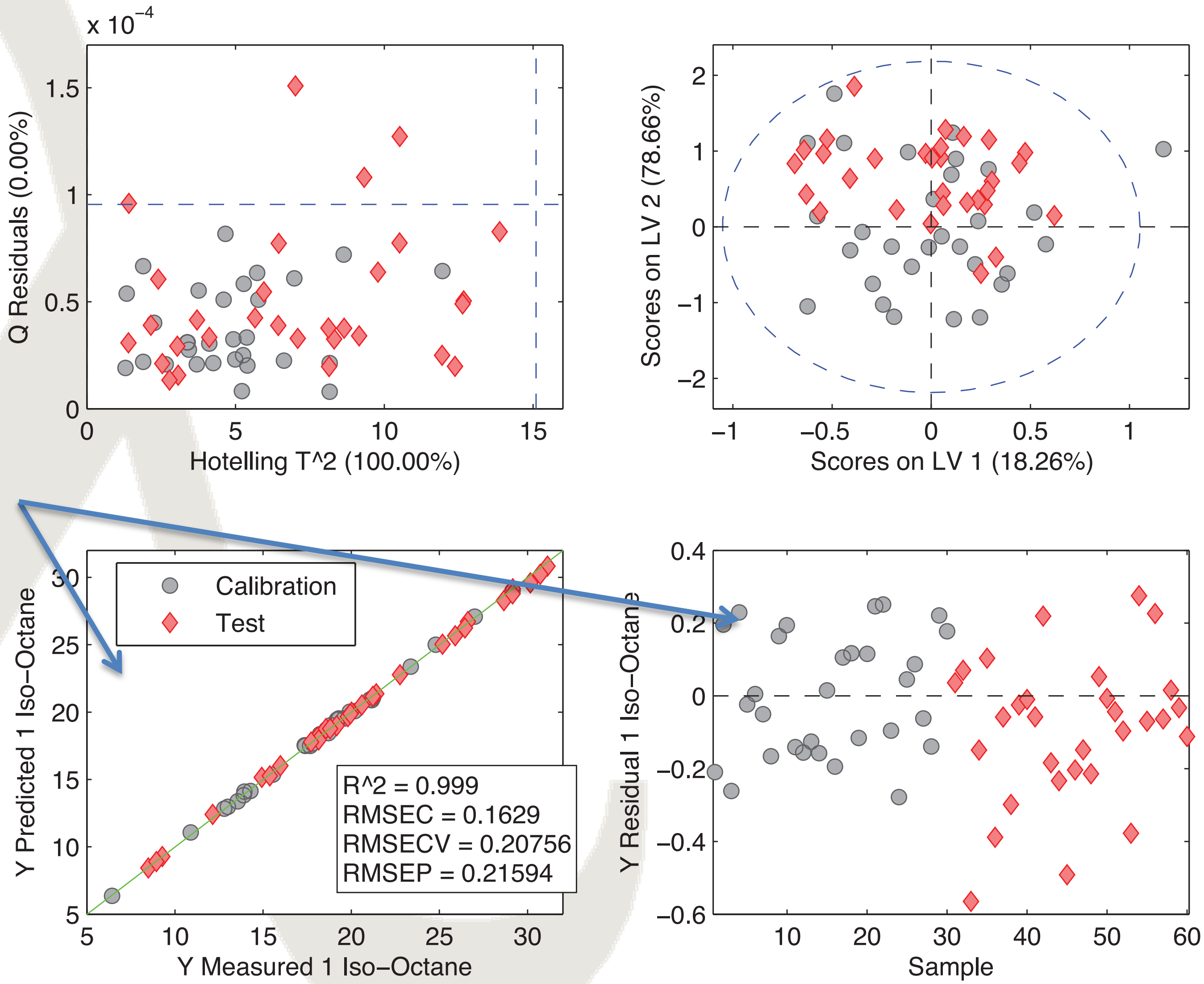
Summary – Q and T^2 Limits

- Need to identify potential failure modes for the process or analytical measurement
  - How do they impact prediction error?
    - How much impact is acceptable?
  - What are the corresponding values for Q and T^2?
  - Choose the most conservative values that will encompass all of the failure modes
- Modeling can be a useful aid in this endeavor
  - More on that later with the model robustness tool

Example of detecting model/data mismatch

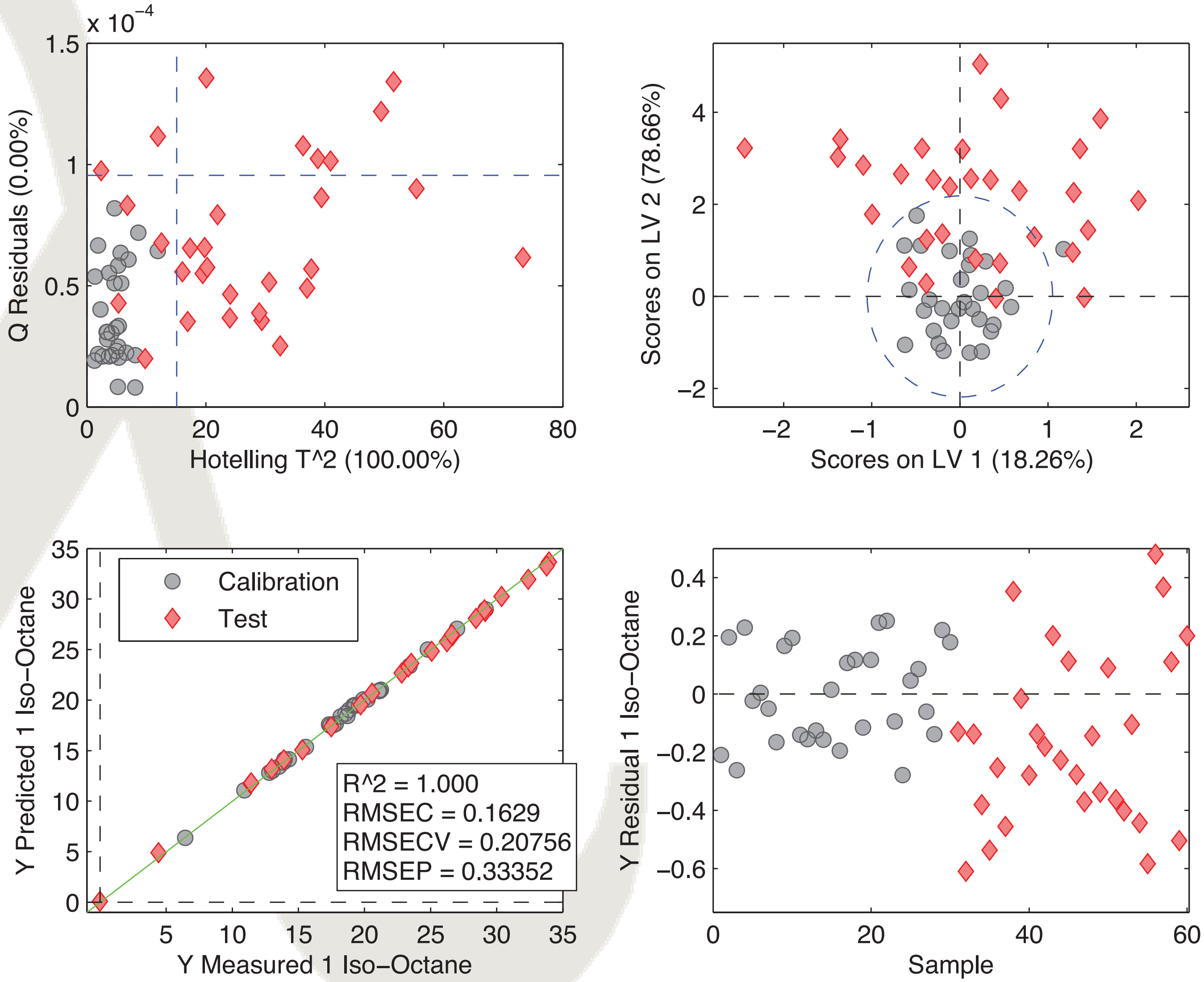
- Semi-synthetic example to illustrate change detection issues
- NIR measurement of iso-octane interferents (heptane, toluene, decane and eventually also xylene)
- CLS model to generate data along with a structured noise model from original data

Normal Operation

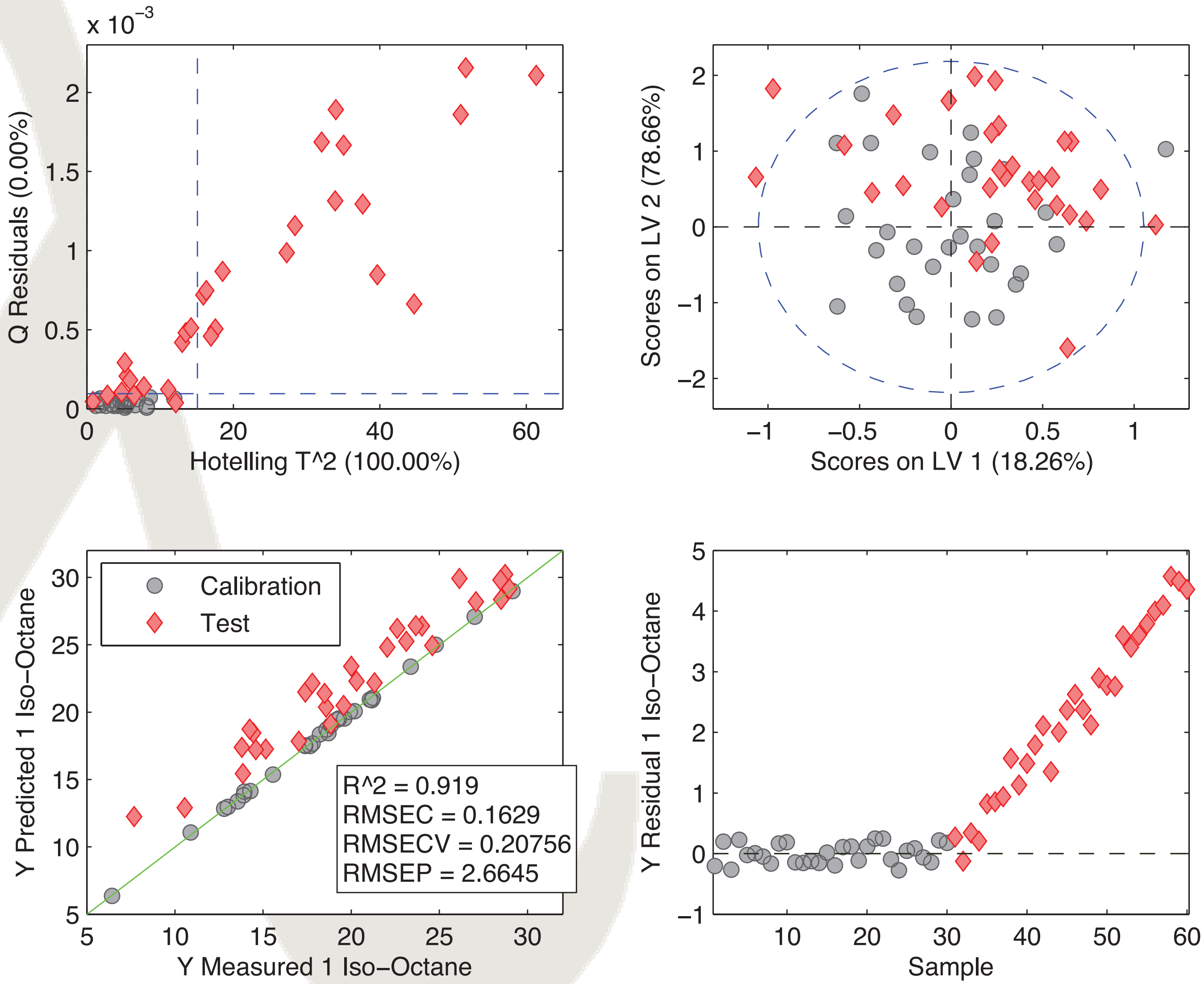


Only available from reference method

Out of Range Samples

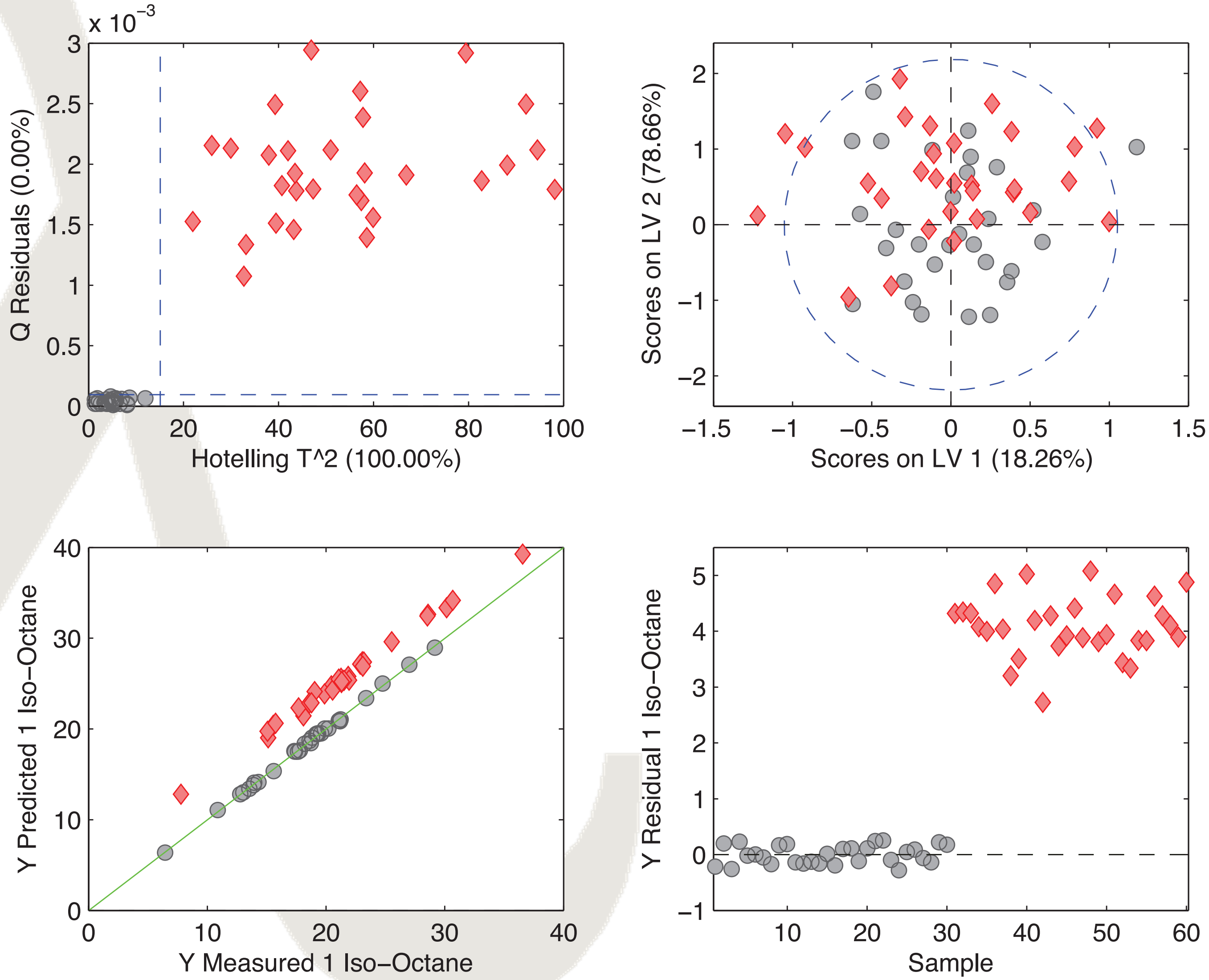


New Interferent

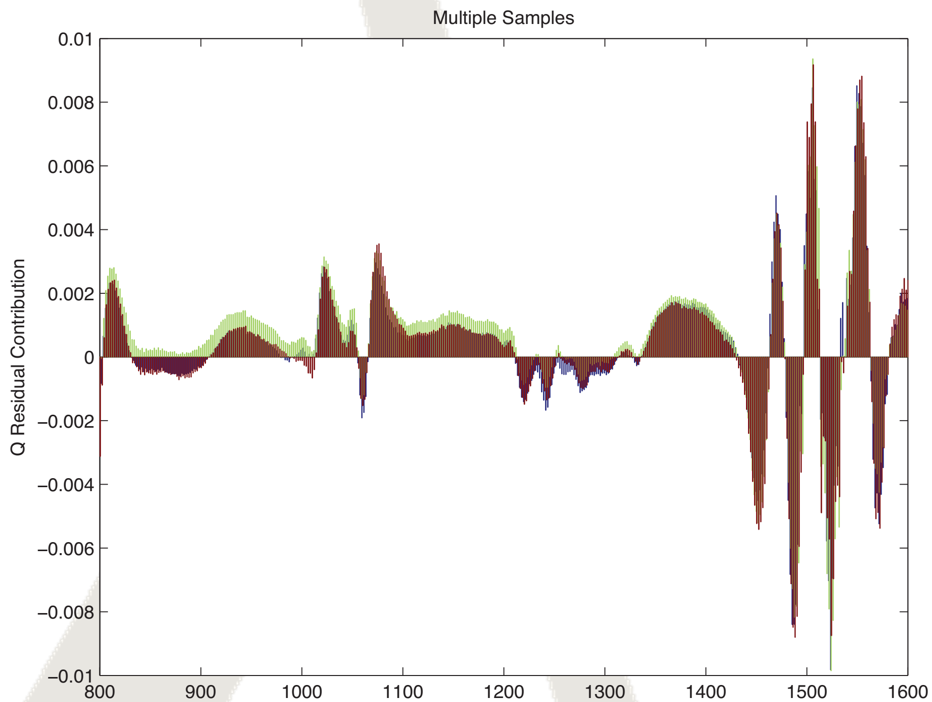


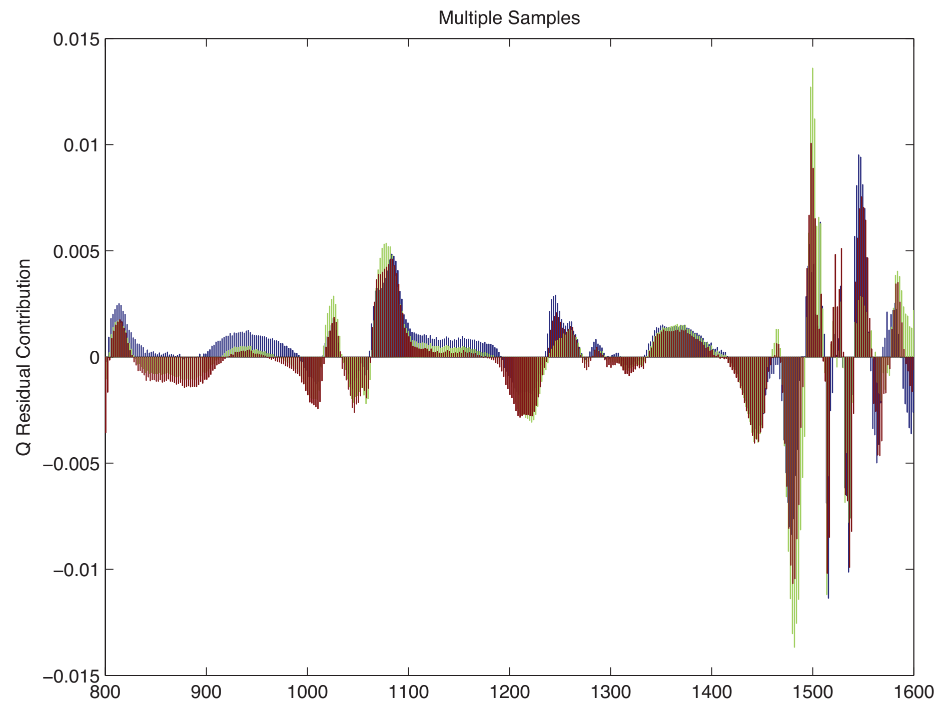
EIGENVECTOR RESEARCH INCORPORATED

Instrument Registration Shift



Residuals





New interferent

Registration shift

Diagnosis Check List

Problem	Sample Distance T^2	Sample Residual Q	Reference agreement	Standards
Out of range	High	Low-Moderate	Good-Fair	Good
New variation	Fair-High	High	Fair-Poor	Good
Relationship change	Fair-Moderate	High	Fair-Poor	Good
Instrument change	Moderate-High	High	Fair-Poor	Poor

Additional Helpful Questions

- What else changed?
  - New raw ingredient, change in upstream equipment, new process setpoint?
- Is shift persistent or variable?
- What does primary method say about interferents?

Determining the cause of the problem

- Might not be obvious from model diagnostics
- Measurement by reference method helpful
- Standard samples can pin down cause unambiguously
- Then what?
  - Expand calibration set
  - Slope and bias correction
  - Instrument standardization

EIGENVECTOR RESEARCH INCORPORATED

Adding Samples to Calibration Set

- Out of range and new interferent problems can usually be solved by adding samples to existing calibration set
- Problem: might take more than a few samples to “balance” calibration
- Solution: upweighting of new samples

Slope and Bias Correction

- Simple to do
- May be appropriate for a constant shift
- Not to be used over and over!
  - Indicates problem is variable interferent or something else
- Automated model updating?

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)
  - Subspace Standardization Transform (SST)
  - Generalized Least Squares Preprocessing (GLS)
  - Orthogonal Signal Correction (OSC)

Standardization Methods

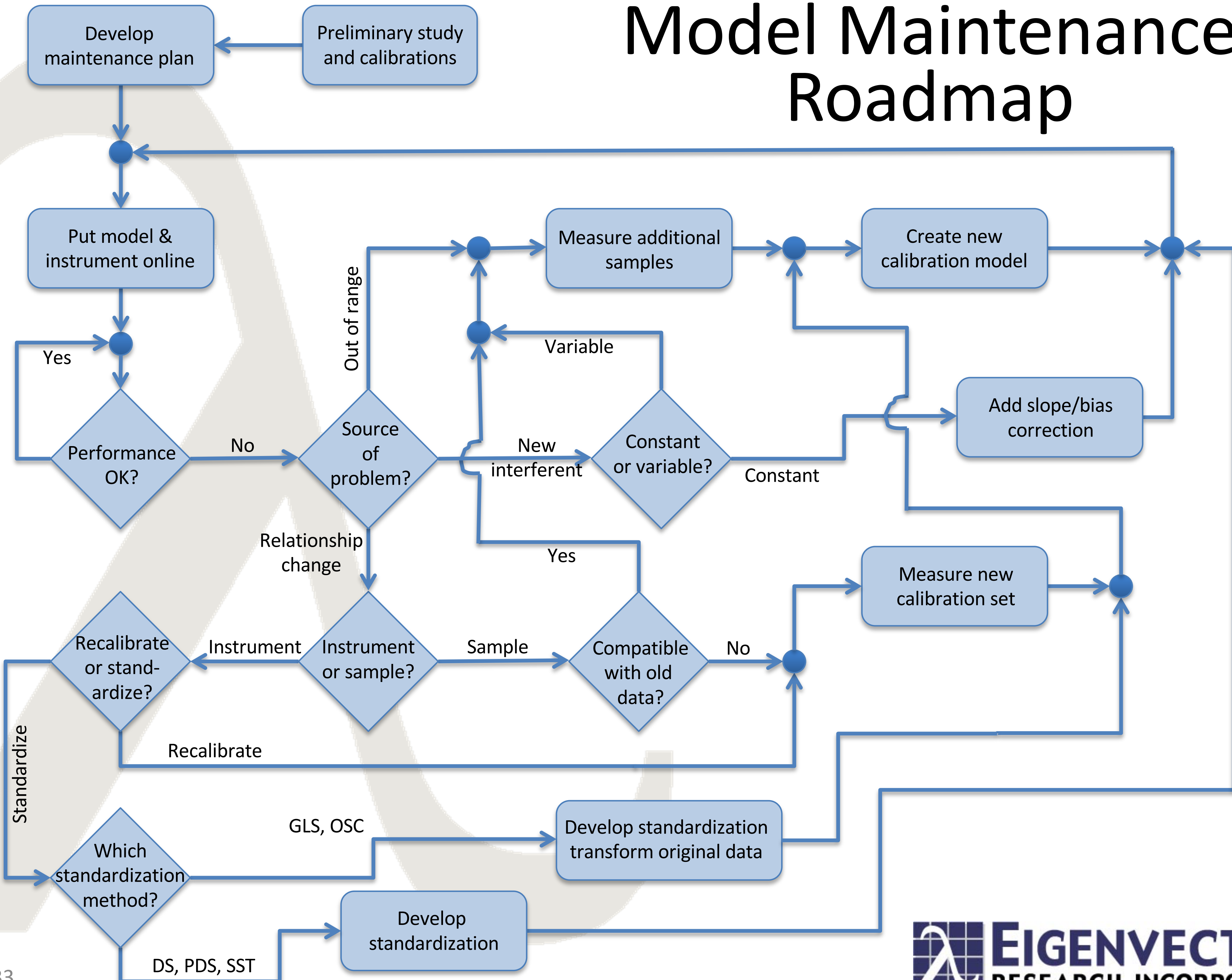
Method	Number of meta-parameters	Y values *not* required?	Use original calibration model?	Spectra un-modified?	Transfer sets *not* function of Y?	Retains net analyte signal?	Can use generic standards?	Number transfer samples required
DS	1							High
PDS	2				✗			Low
SST	1				✗			Medium
GLS	1		✗	✗		✗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

EIGENVECTOR RESEARCH INCORPORATED

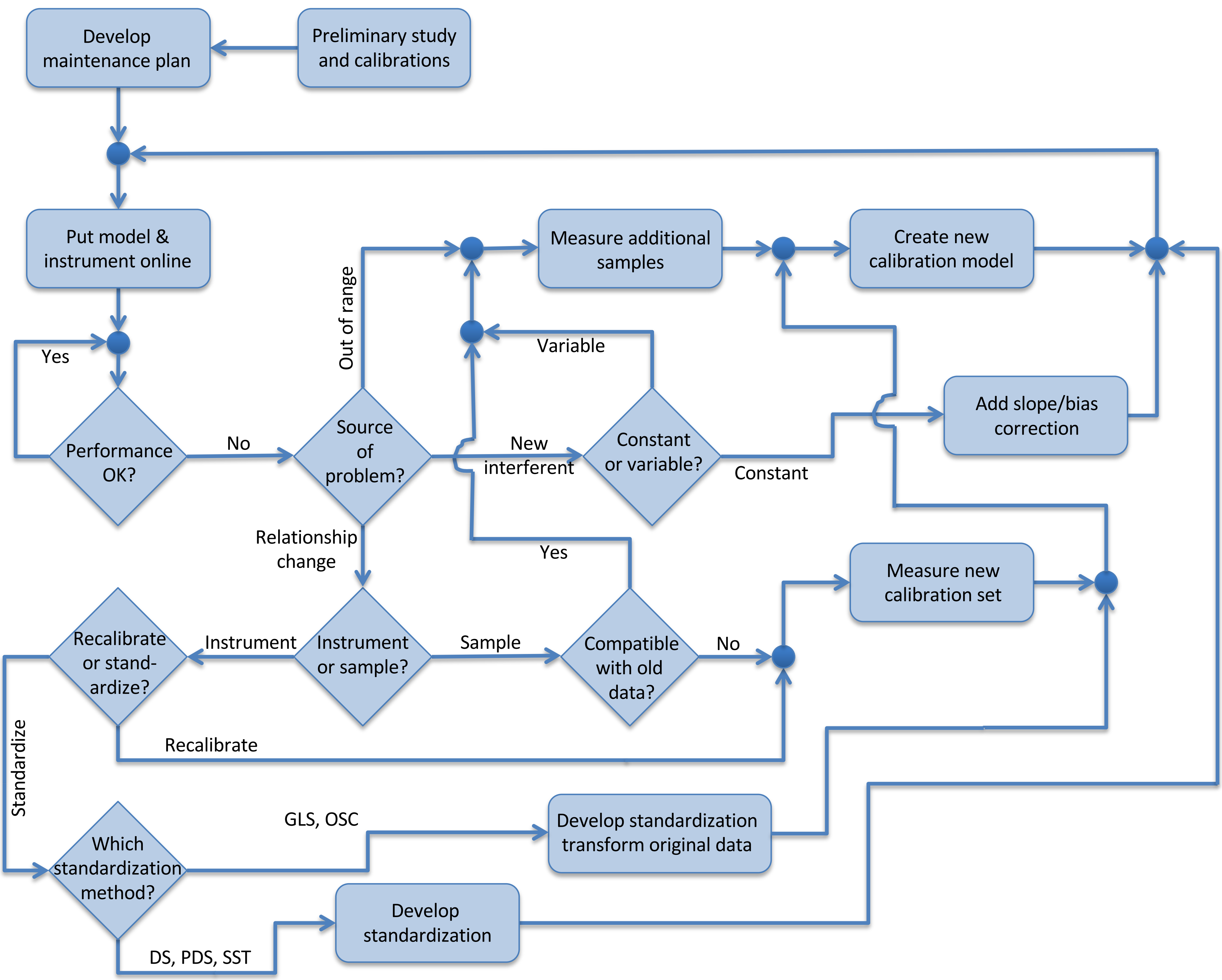
Putting it all together

- Have covered the pieces
  - Detecting change and performance degradation
  - Identifying the problem
  - Methods for updating/correcting models
  - Standardization methods
- How does this all fit together?

Model Maintenance Roadmap



EIGENVECTOR RESEARCH INCORPORATED


Develop maintenance plan
Preliminary study and calibrations
Put model & instrument online
Yes
Performance OK?
No
Out of range
Source of problem?
Measure additional samples
Create new calibration model
Variable
New interferent
Constant or variable?
Constant
Add slope/bias correction
Relationship change
Yes
Measure new calibration set
Recalibrate or standardize?
Instrument
Instrument or sample?
Sample
Compatible with old data?
No
Recalibrate
Standardize
GLS, OSC
Develop standardization transform original data
Which standardization method?
Develop standardization
DS, PDS, SST

Avoiding Model Maintenance

- Some models more robust to new analytes and changes in data than others
- Highly dependent on preprocessing options and number of factors in models

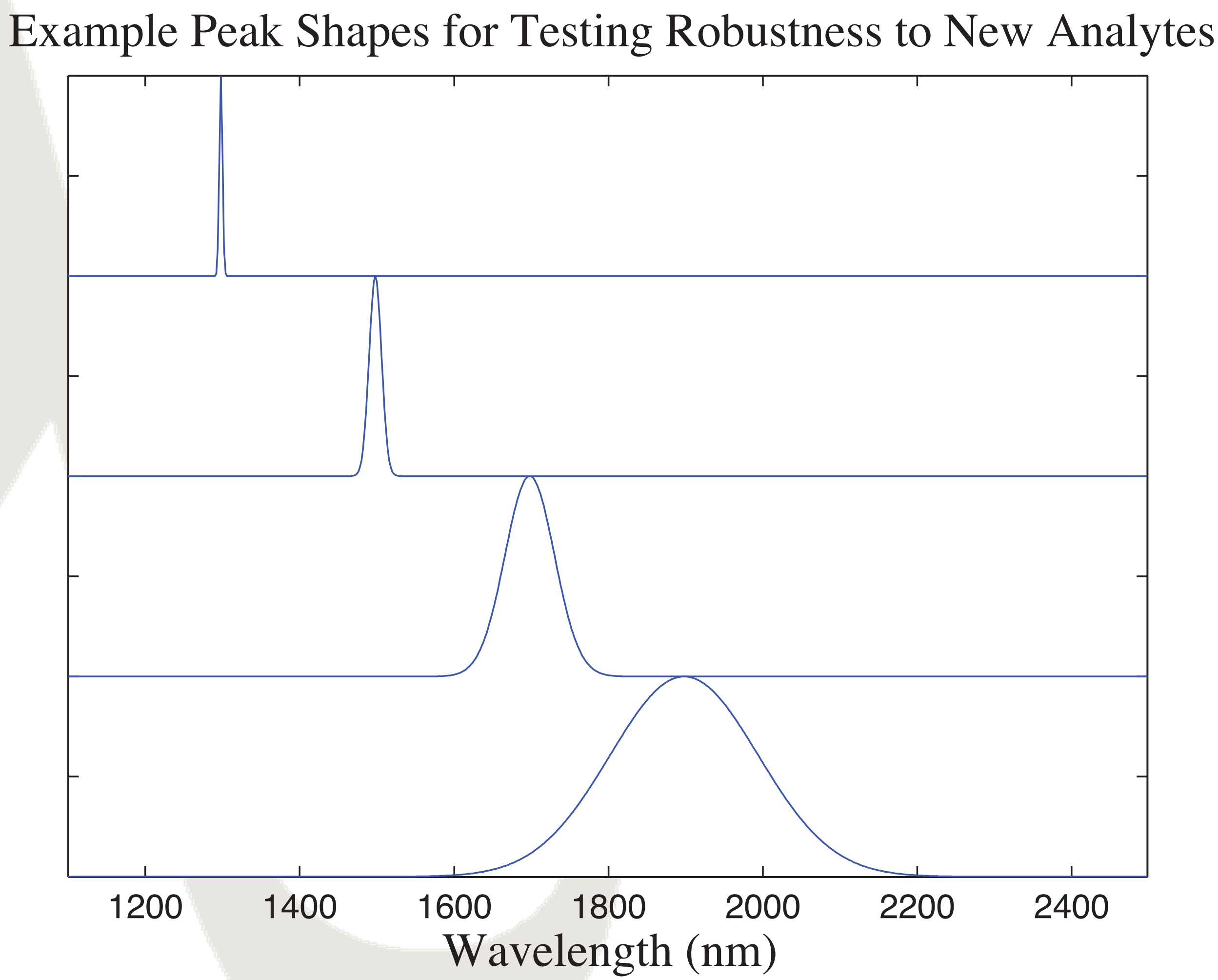
Robustness Tests

- Series of functions developed to test model against system changes
  - Develop model with desired preprocessing, #LVs, etc.
  - "Perturb" test data set
  - Apply calibration model to "perturbed" data
  - Look at prediction error as function of perturbations
  - Test and compare multiple models

Perturbations

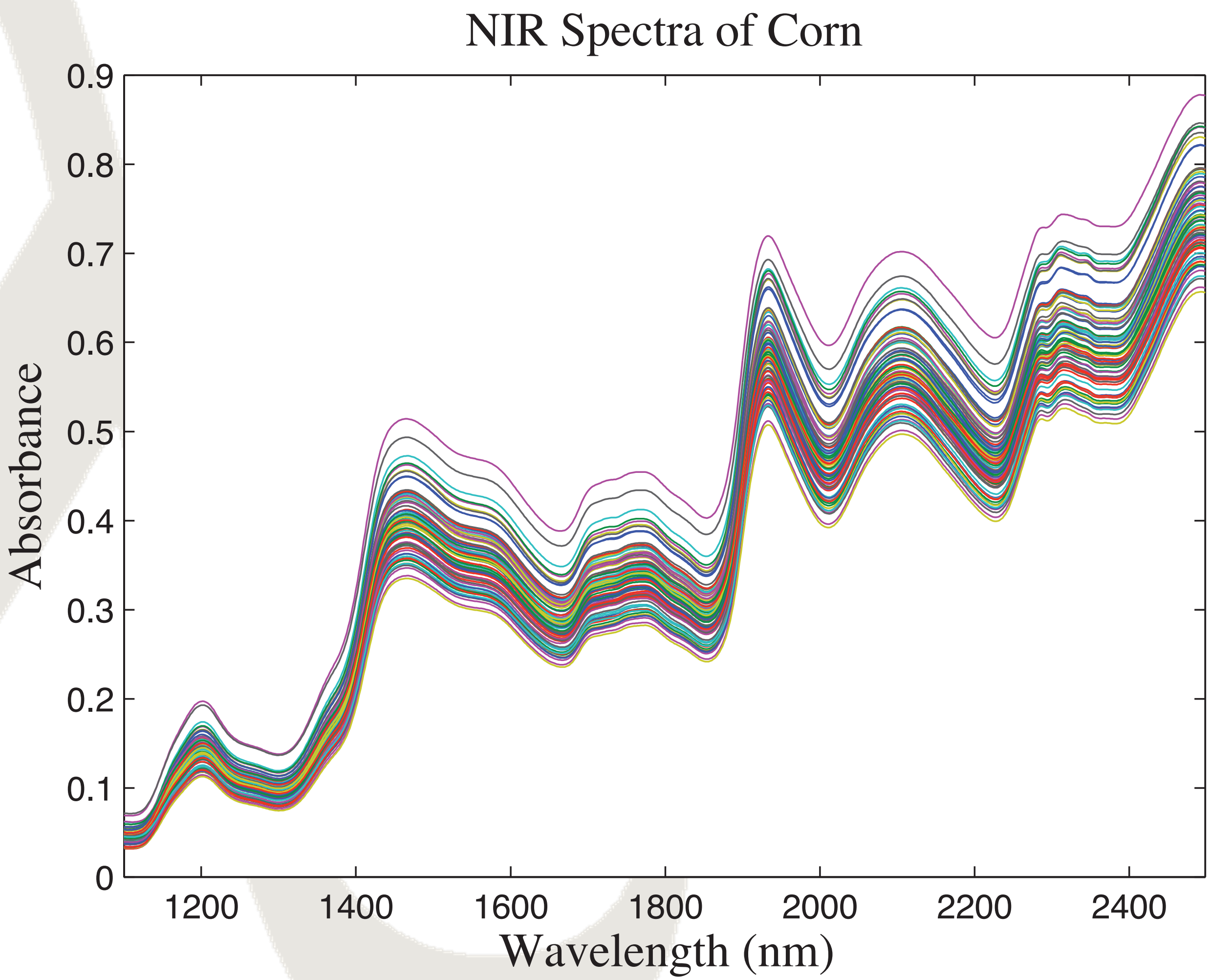
- New analyte – add Gaussian peak of variable width across wavelength range
- Wavelength registration shift – shift spectra left-right as well as expand and contract
- Others:
  - Baseline shift – change offset and slope
  - Stray light – add fraction of signal before log transform
  - Temperature – decrease resolution and vary path length
  - Noise variation – add noise with varying bandwidth

New Analyte



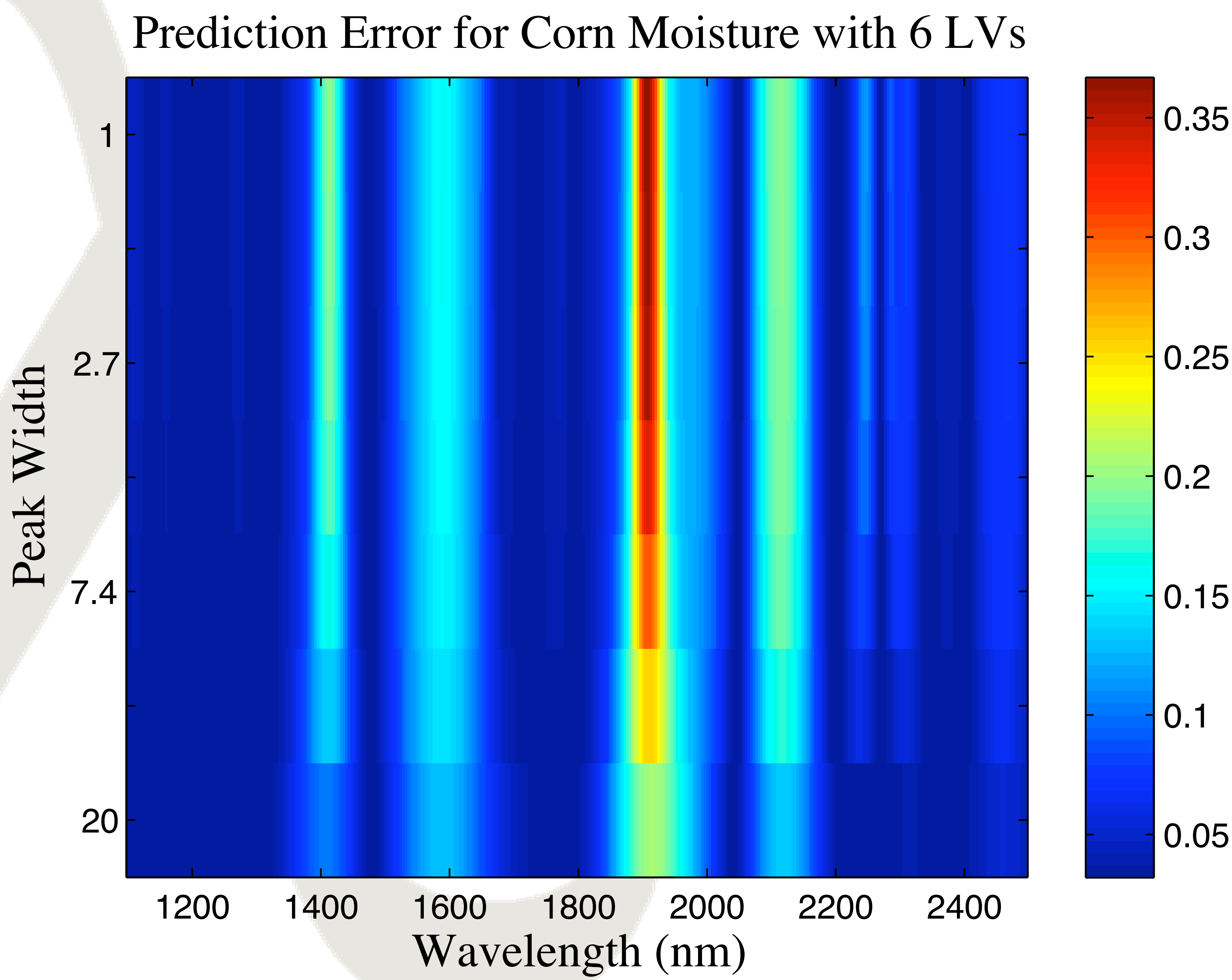
Example Peak Shapes for Testing Robustness to New Analytes
1200
1400
1600
1800
2000
2200
2400
Wavelength (nm)

Example Data: Corn NIR

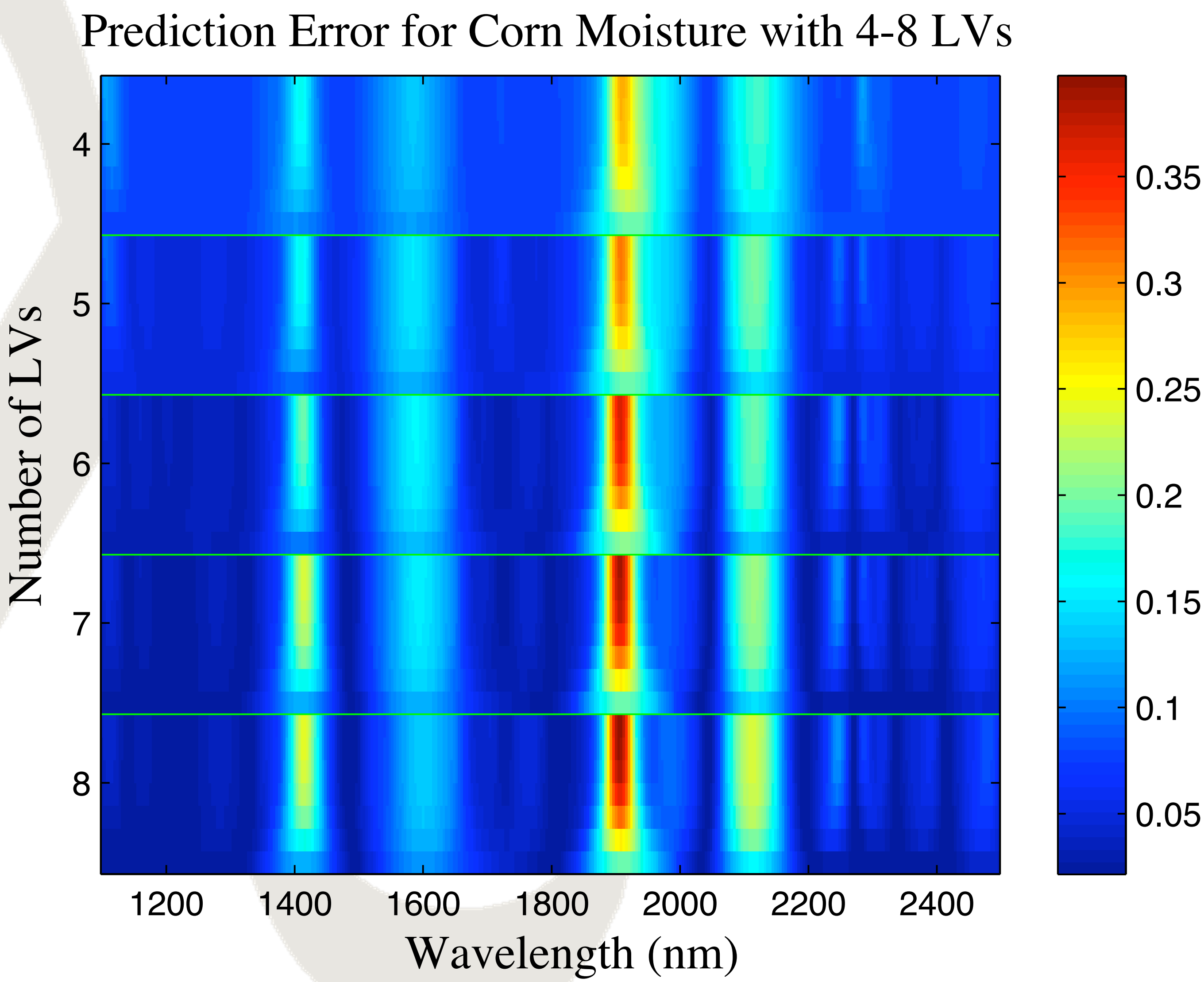


EIGENVECTOR RESEARCH INCORPORATED

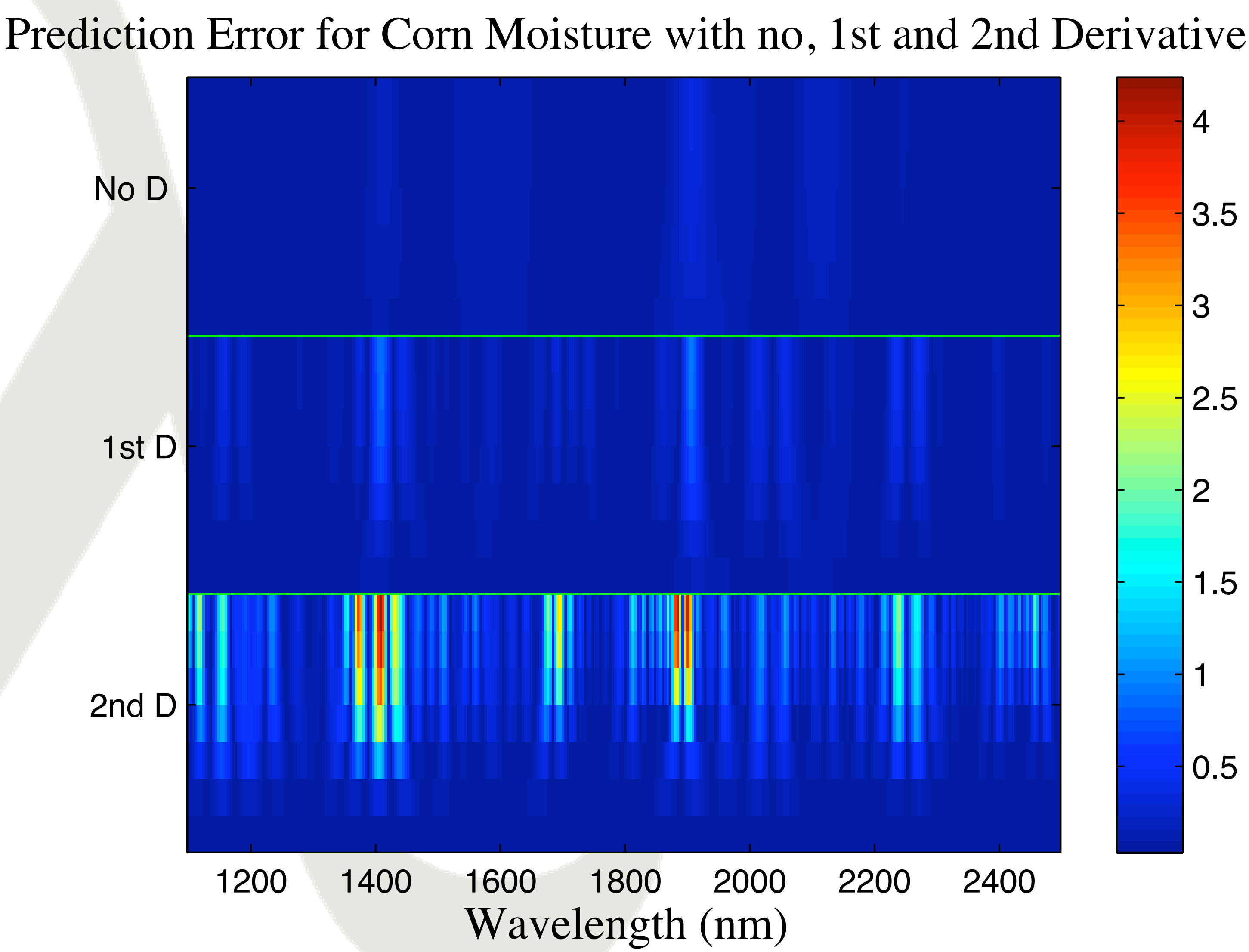
Test Corn Model



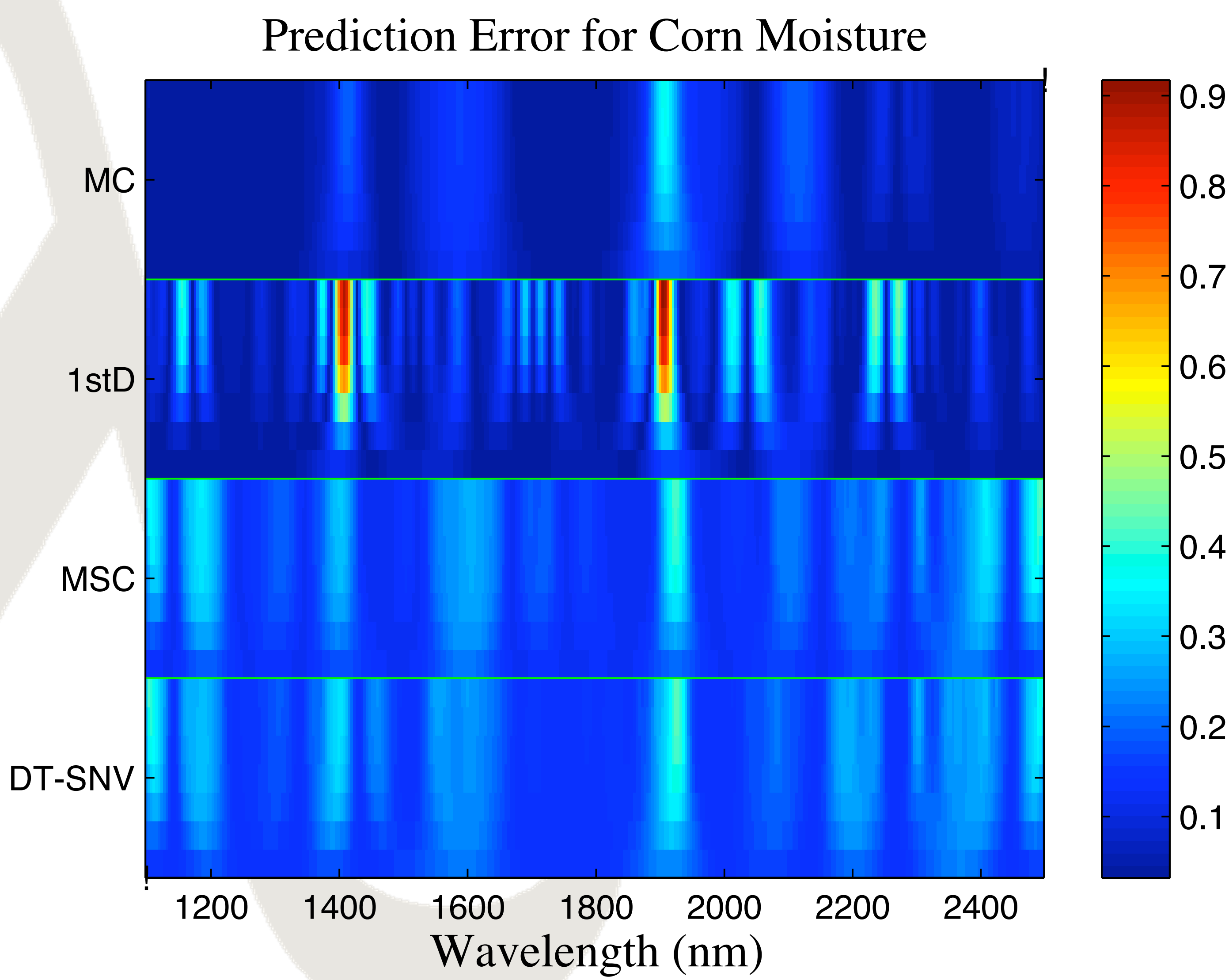
Compare Models- #LVs



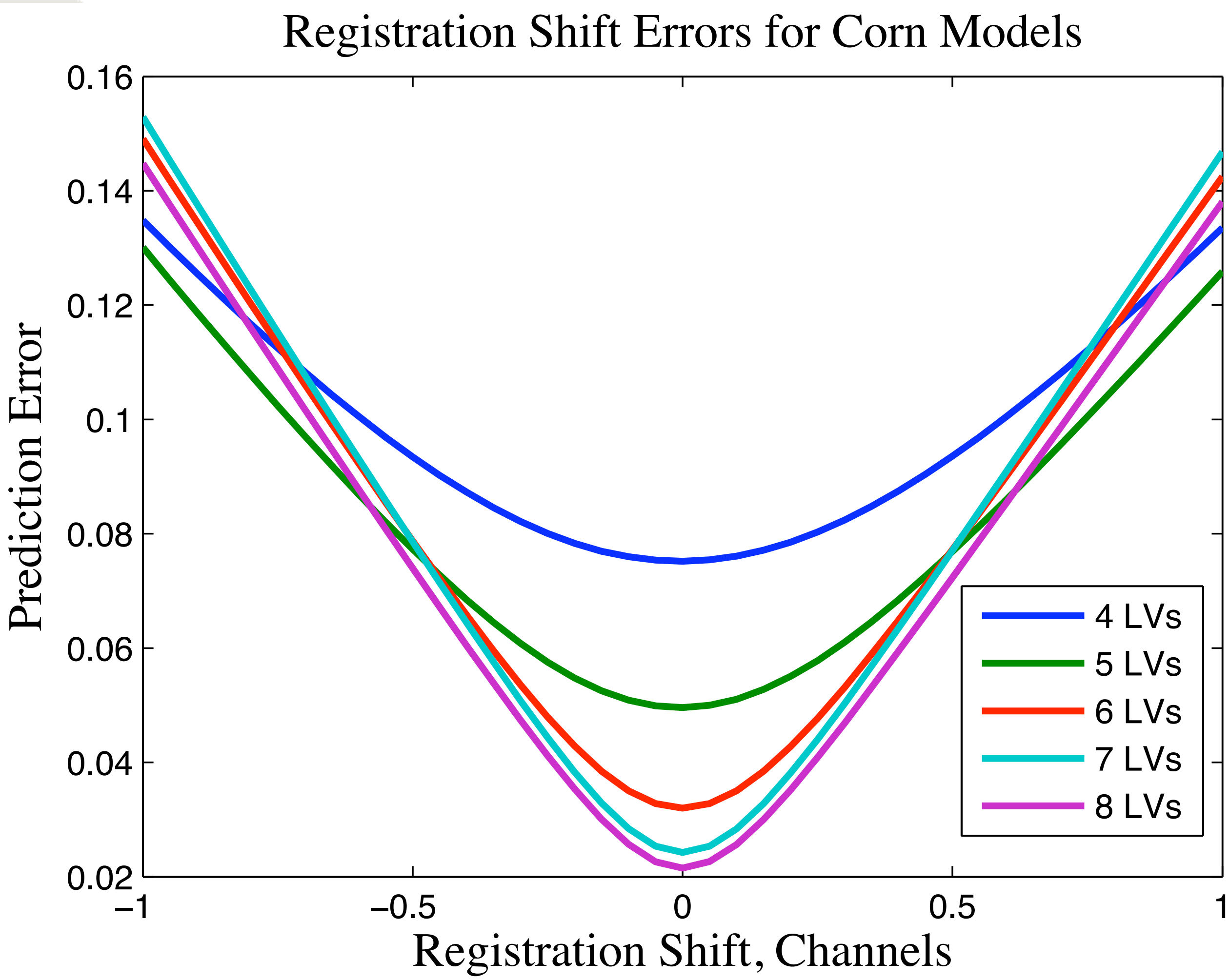
Compare Models-Derivative



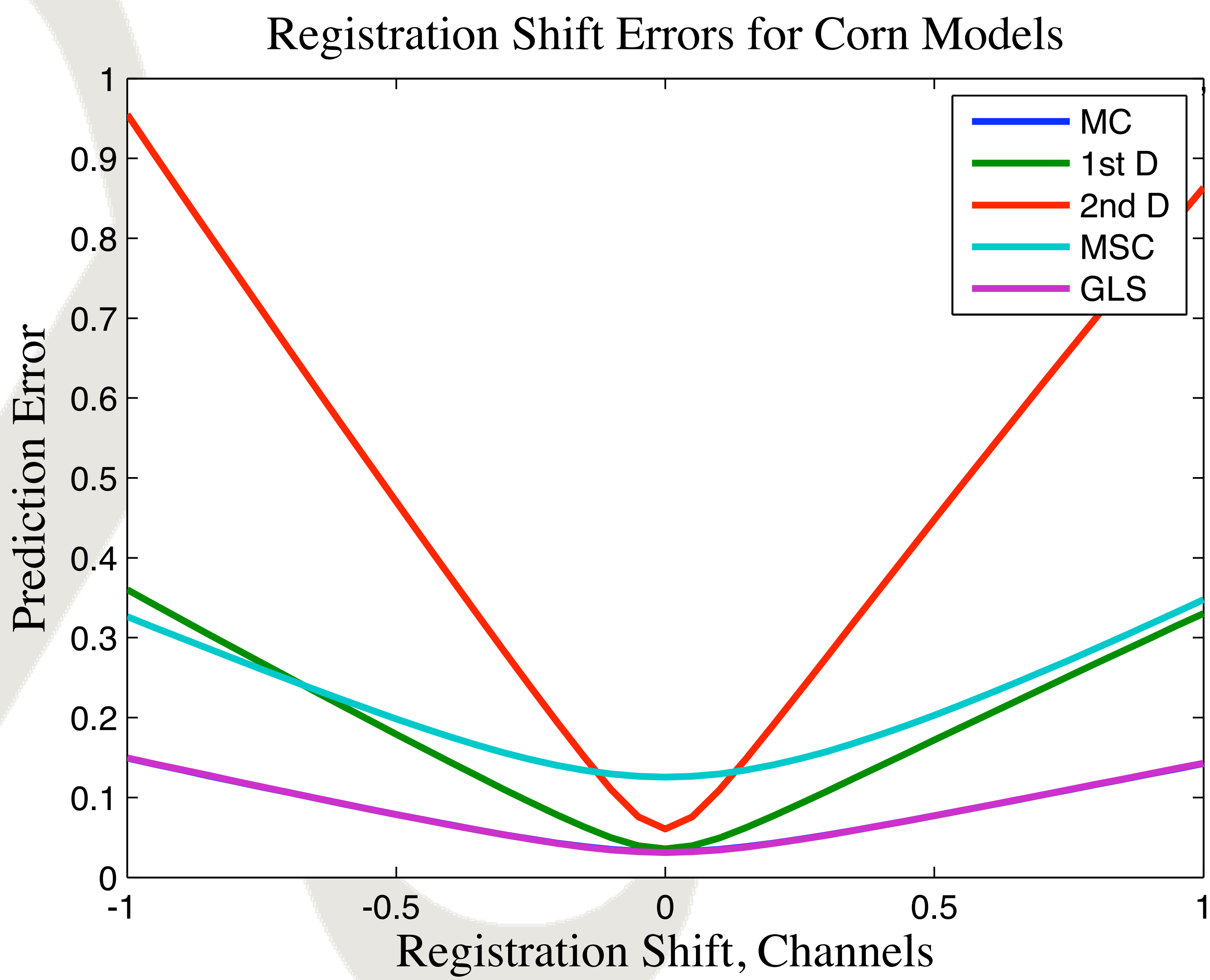
Other Preprocessing



Shift with #LVs



Shift with Preprocessing



Conclusions

- Plan and budget (!) for model maintenance
- Many elements
  - Performance monitoring
  - Problem detection and identification
  - Standardization protocols
  - Remodeling and revalidating
- The “Road Map” can be customized for specific applications
- Model robustness testing can help minimize the need for model updating

EIGENVECTOR RESEARCH INCORPORATED