Model Maintenance: When, how and how to avoid it

Barry M. Wise and Robert T. Roginski

Eigenvector Research Inc., Manson, WA, USA

Outline

- Introduction to Model Maintenance
- Identifying the need for action
- Model updating methods
- Instrument standardization methods
- Avoiding model maintenance
- Conclusions

B.M. Wise and R.T. Roginski, "Model Maintenance: the unrecognized cost in PAT and QbD," **Chemistry Today**, Vol. 33(2), pps. 38-43, March/April, 2015.
B.M. Wise and R.T. Roginski, "A Calibration Model Maintenance Roadmap," **Proceedings of ADCHEM 2015**, Whistler, BC, Canada, pps. 260-265, June, 2015.

Definition and Goal

- 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?

Diagnostics Limits

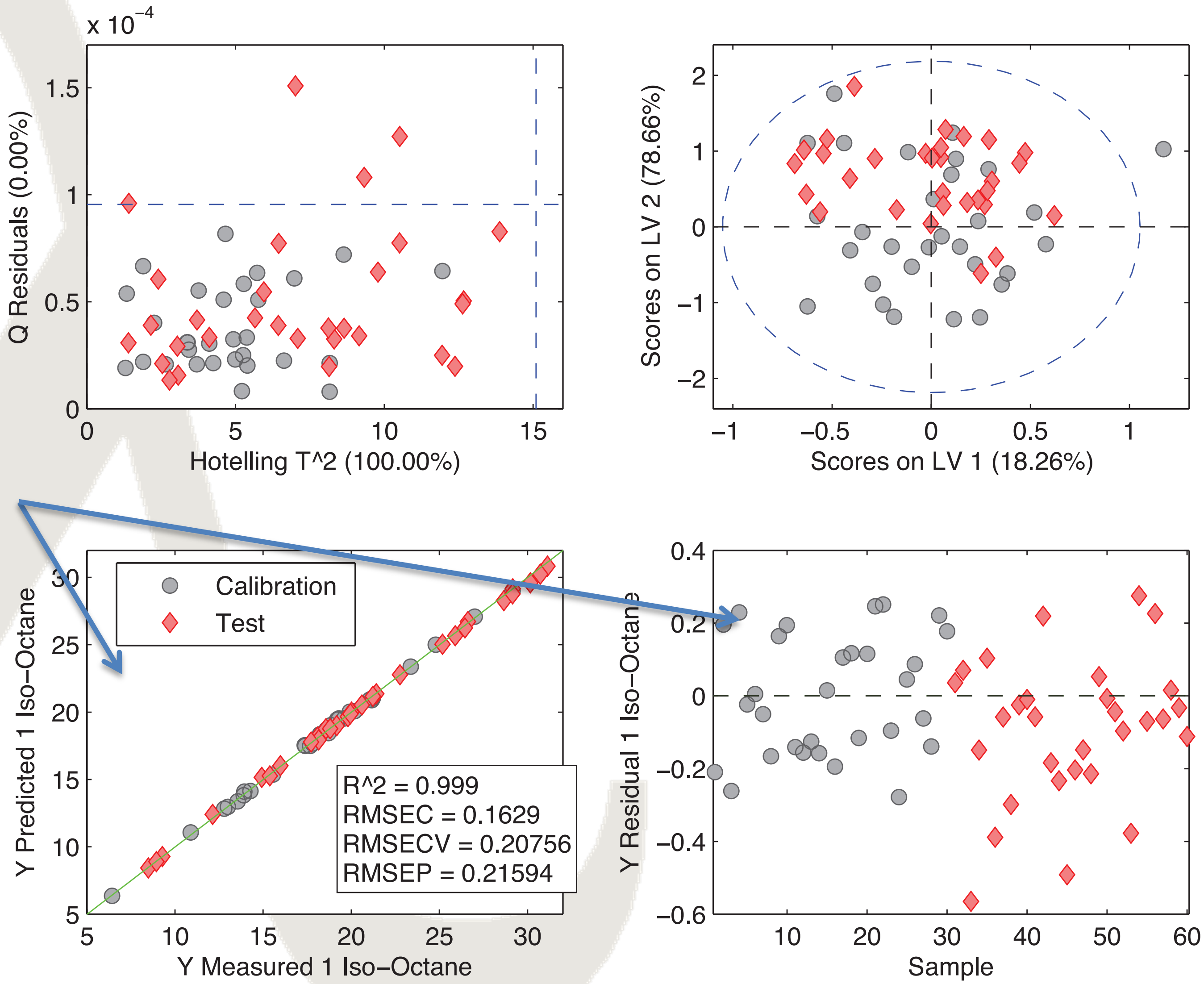
- By default, limits for Q and T^2 are generally provided based on some type of confidence limit
  - These limit values are statistically based
- Meaningful limits require knowledge of the process or measurement
  - Observations are needed that represent
    - “bad” states
    - out of spec product
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

*these may be challenging to obtain

Example of detecting model/data mismatch

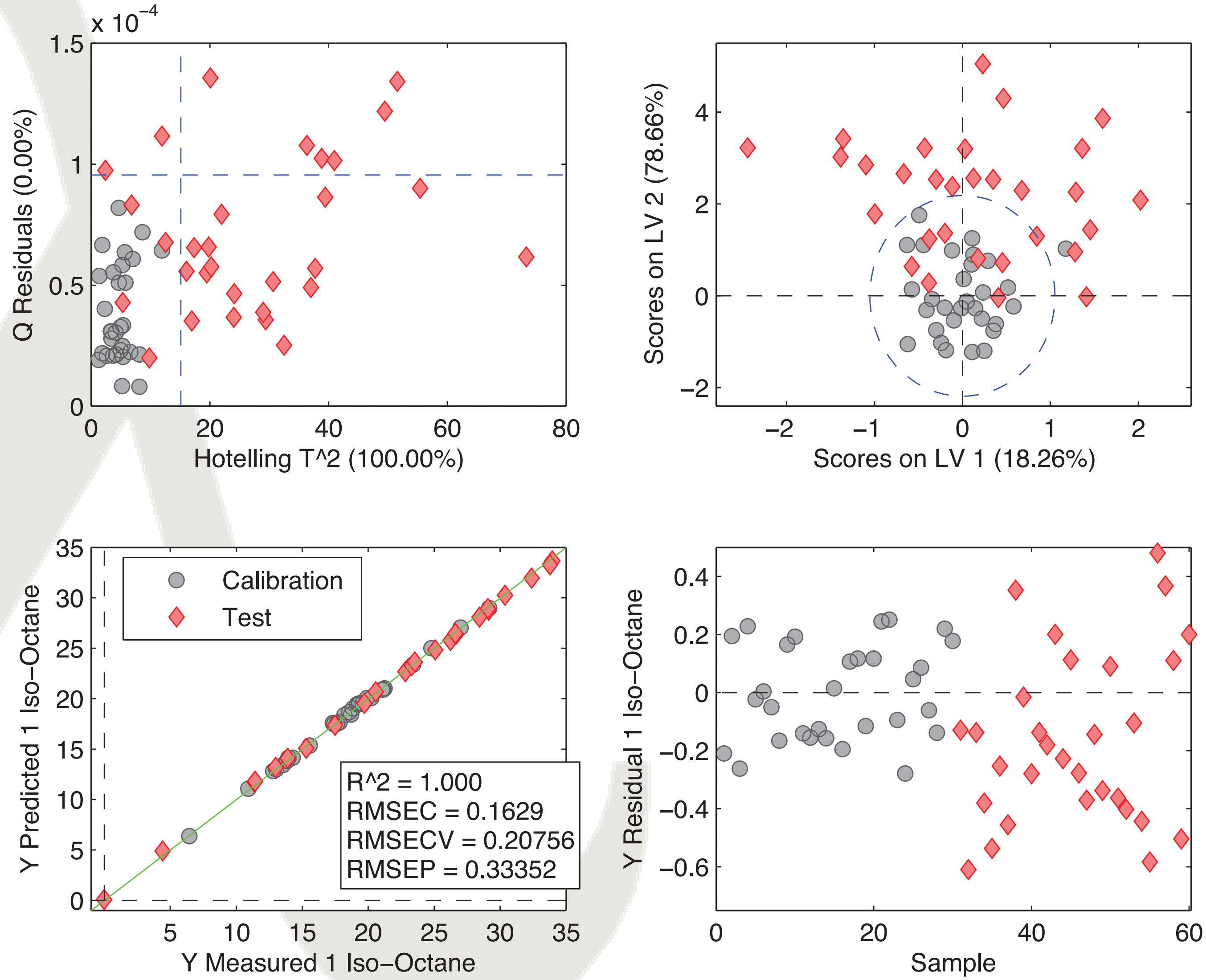
- Semi-synthetic example to illustrate change detection issues
- NIR measurement of iso-octane interferents (heptane, toluene, decane and eventually also xylene)
- CLS model to generate data along with a structured noise model from original data

Normal Operation

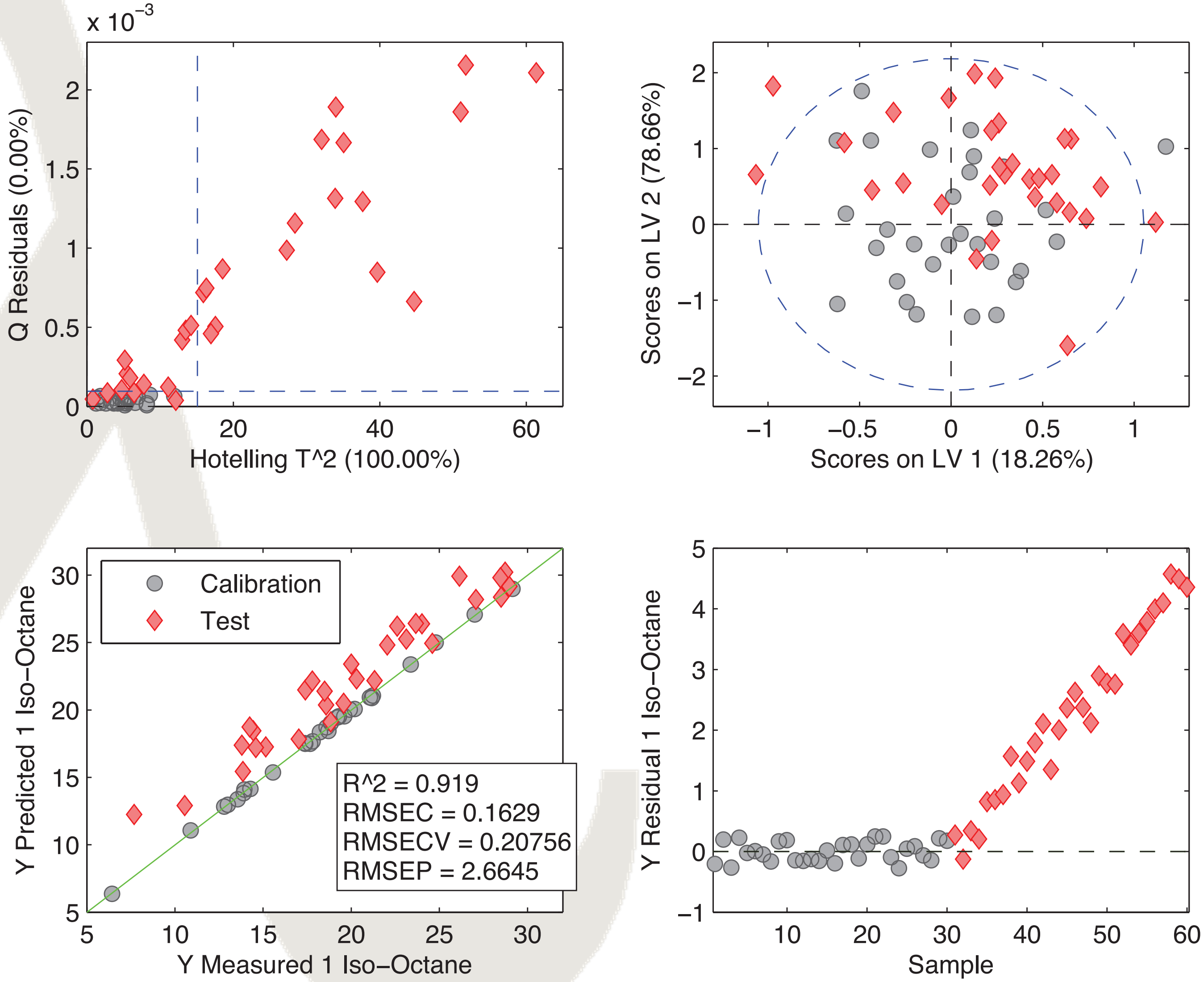


Only available from reference method

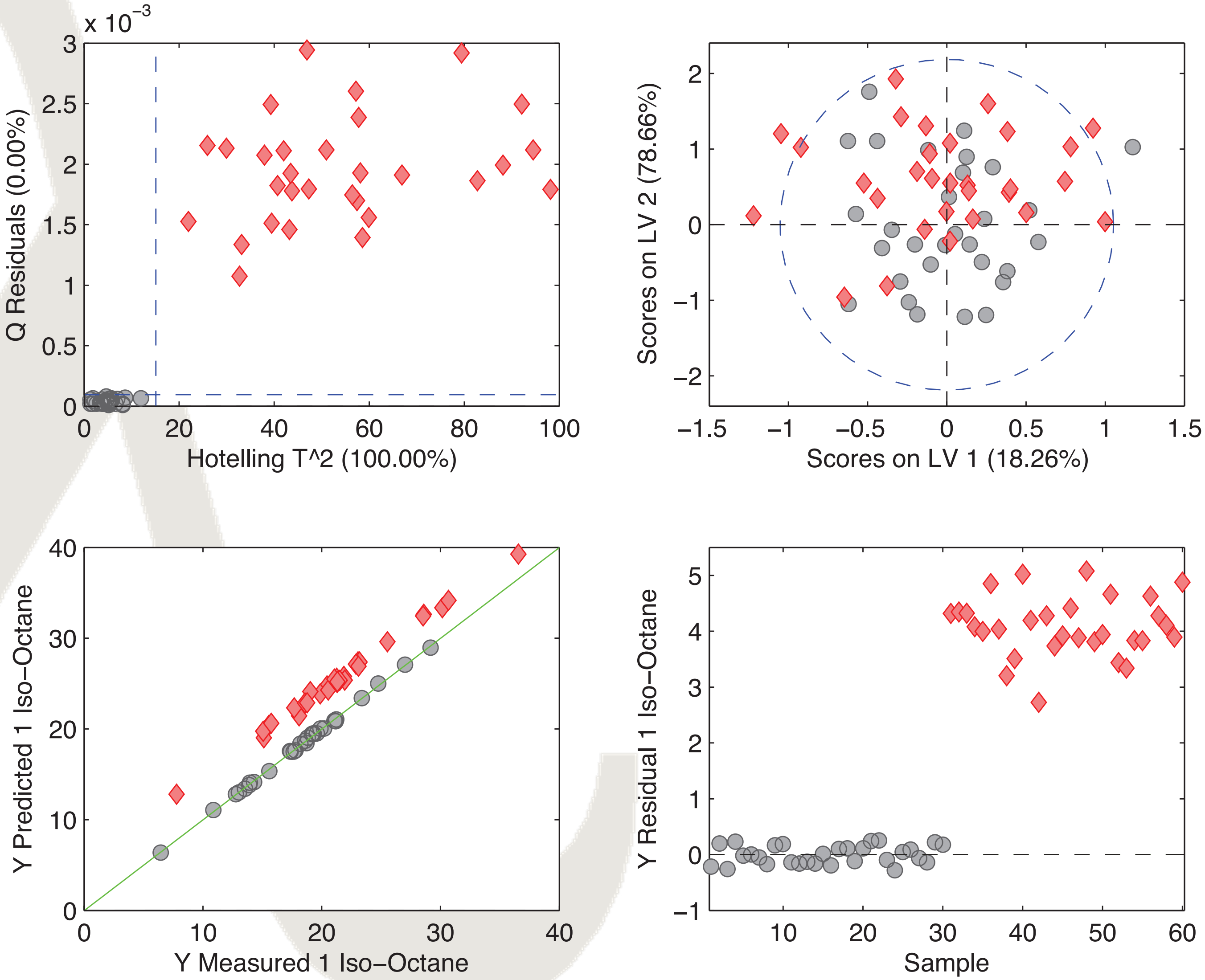
Out of Range Samples



New Interferent



Instrument Registration Shift



EIGENVECTOR RESEARCH INCORPORATED

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

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

EIGENVECTOR RESEARCH INCORPORATED

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?

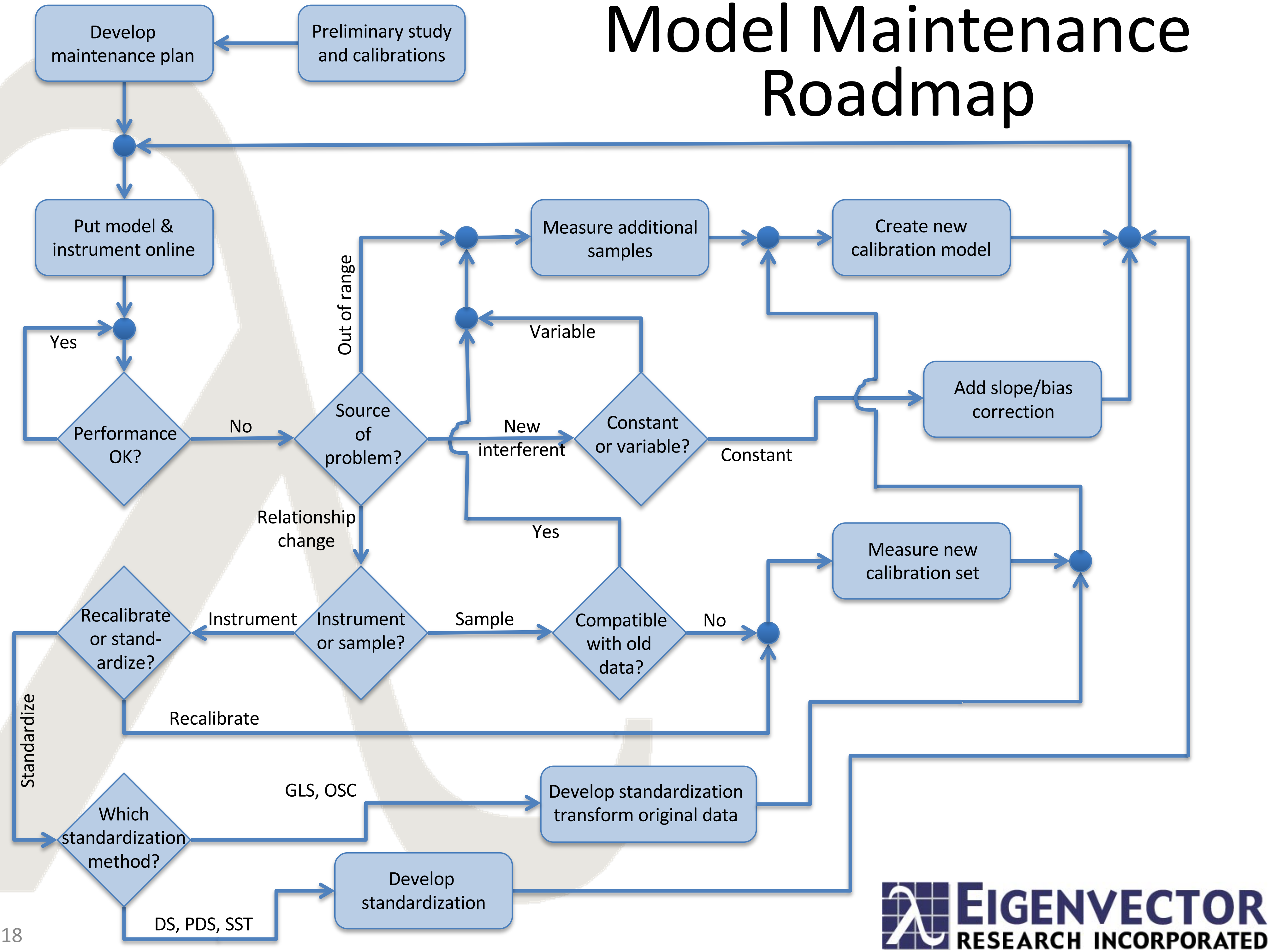
Standardization Methods

- 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)

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



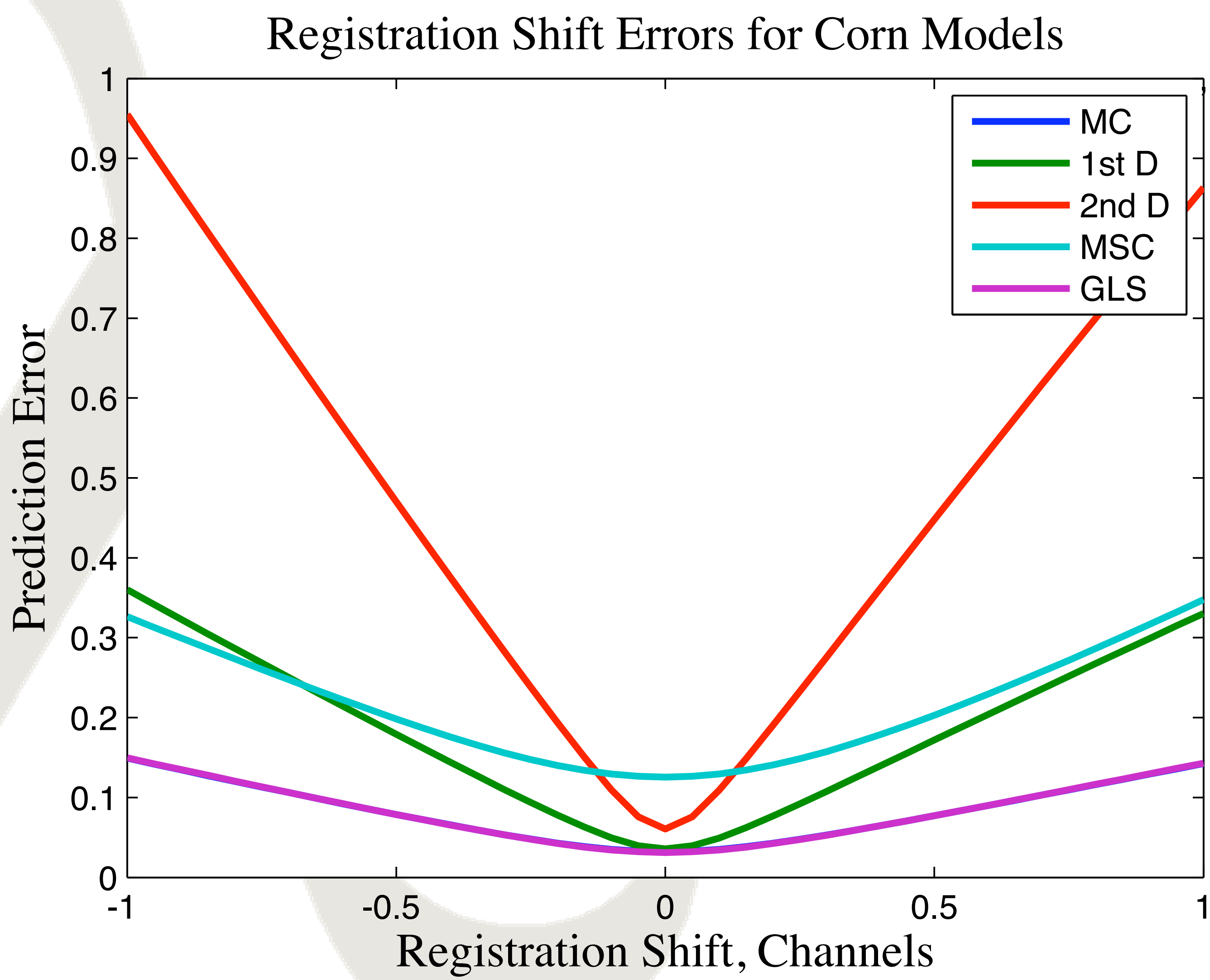
Avoiding Model Maintenance

- 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

Shift with Preprocessing



Conclusions

- Plan and budget (!) for model maintenance
- Many elements
  - Performance monitoring
  - Problem detection and identification
  - Standardization protocols
  - Remodeling and revalidating
- Model robustness testing can help minimize the need for model updating