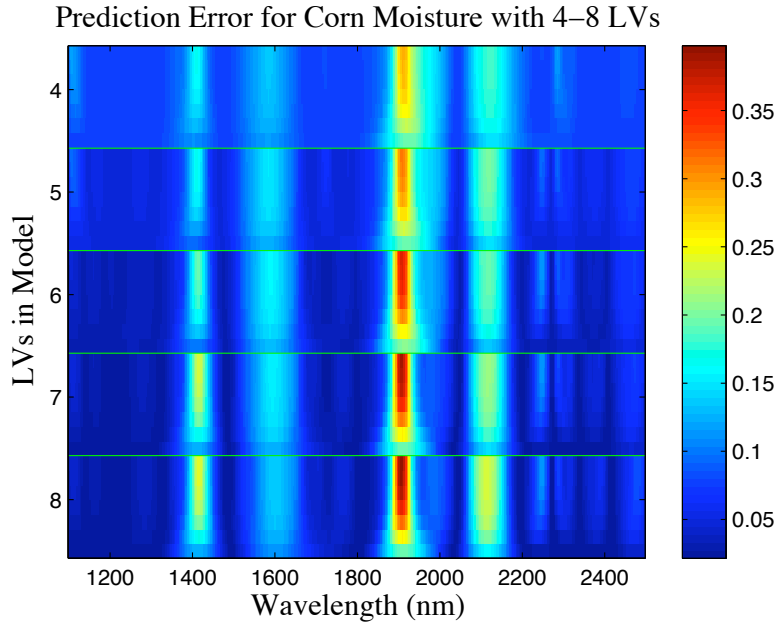


Tools for Multivariate Calibration Robustness Testing with Observations on Effects of Data Preprocessing

Barry M. Wise
EIGENVECTOR RESEARCH INCORPORATED

Introduction

- When developing calibration models focus is generally on improving prediction error
- Models often developed with small amount of data taken over relatively short time
- Prediction errors over long term often dominated by artifacts not represented in calibration data
  - Spectrometer
  - Sample

EIGENVECTOR RESEARCH INCORPORATED

What Constitutes a Good Model?

- Acceptable prediction error
  - Note: *not* best achievable
- Longevity, *i.e.* robustness to minor changes

Possible Changes to System

- Sample
  - New analyte(s)
  - Changes in physical properties (e.g. scattering)
  - Temperature
  - Pressure
- Instrument
  - Wavelength registration shift
  - Stray light
  - Resolution
  - Noise

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

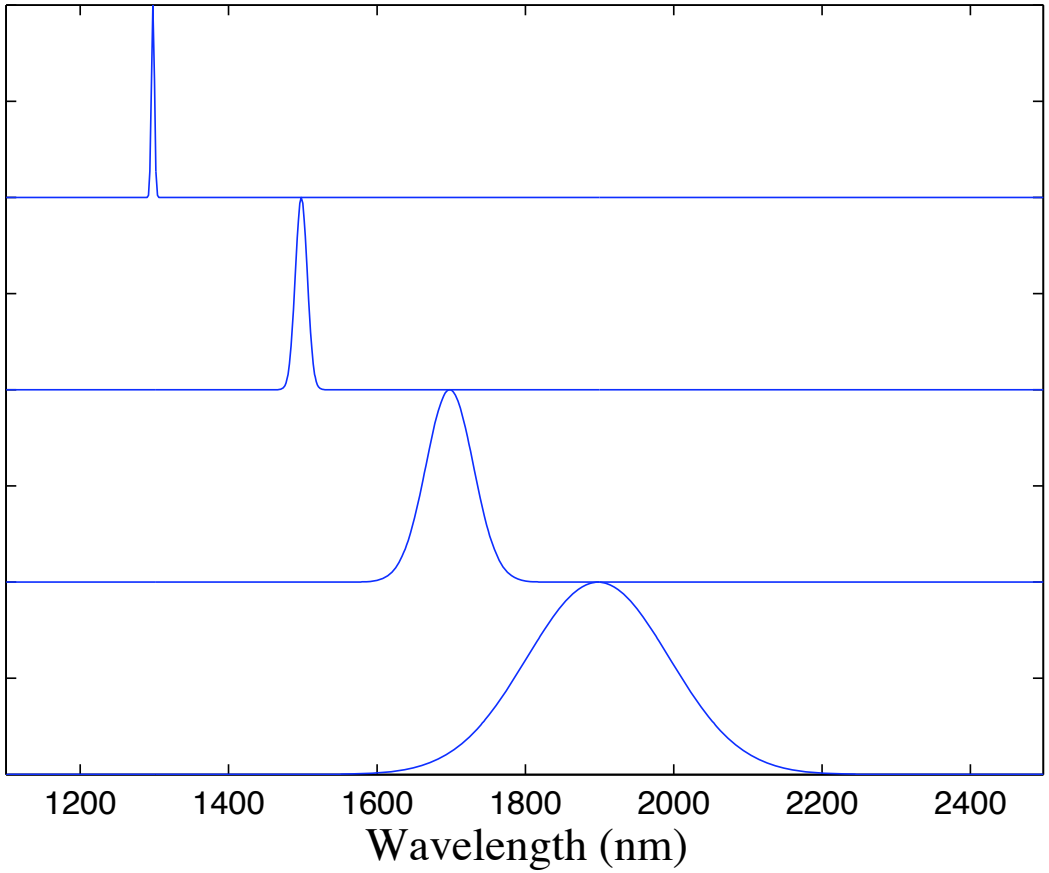
Perturbations

- New analyte – add Gaussian peak of variable width across wavelength range
- Wavelength registration shift – shift spectra left-right as well as expand and contract
- 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

EIGENVECTOR RESEARCH INCORPORATED

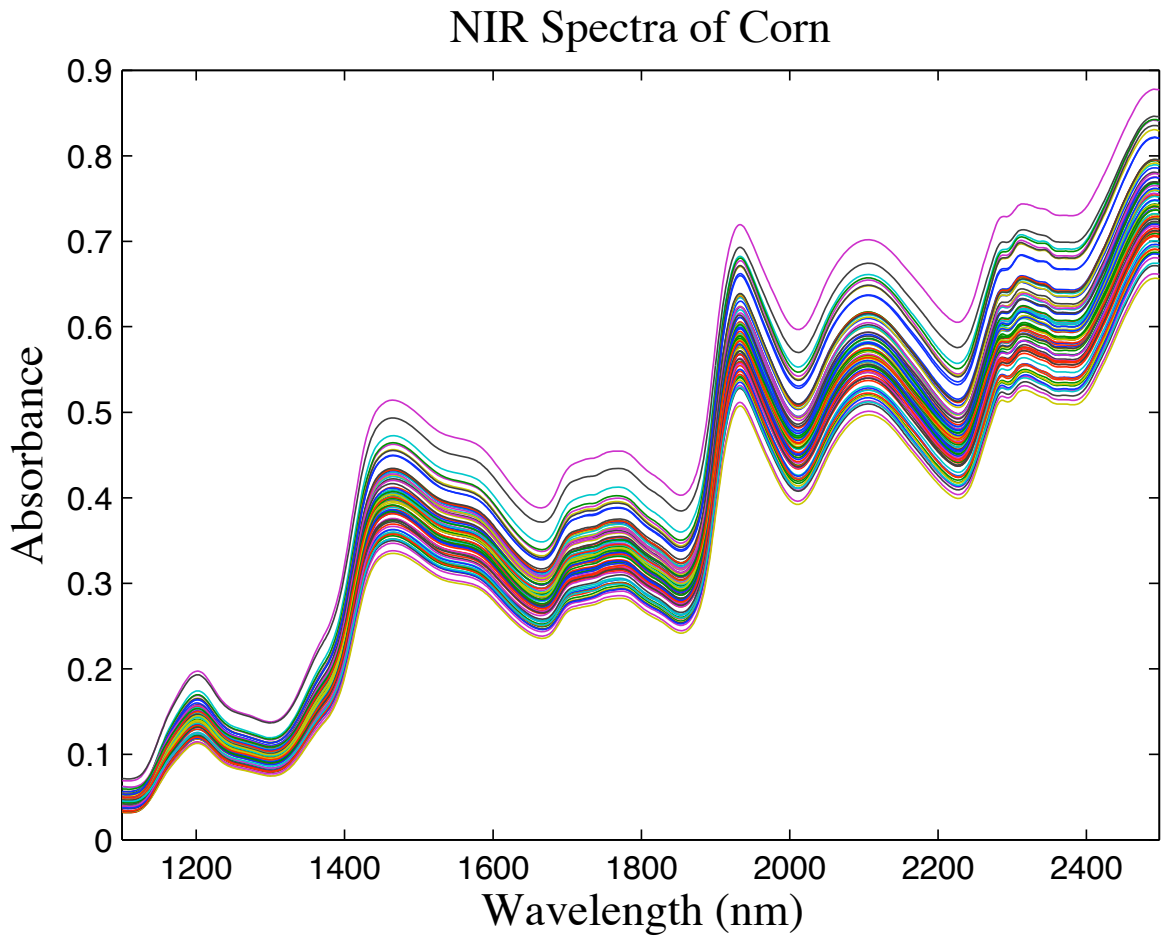
New Analyte

Example Peak Shapes for Testing Robustness to New Analytes



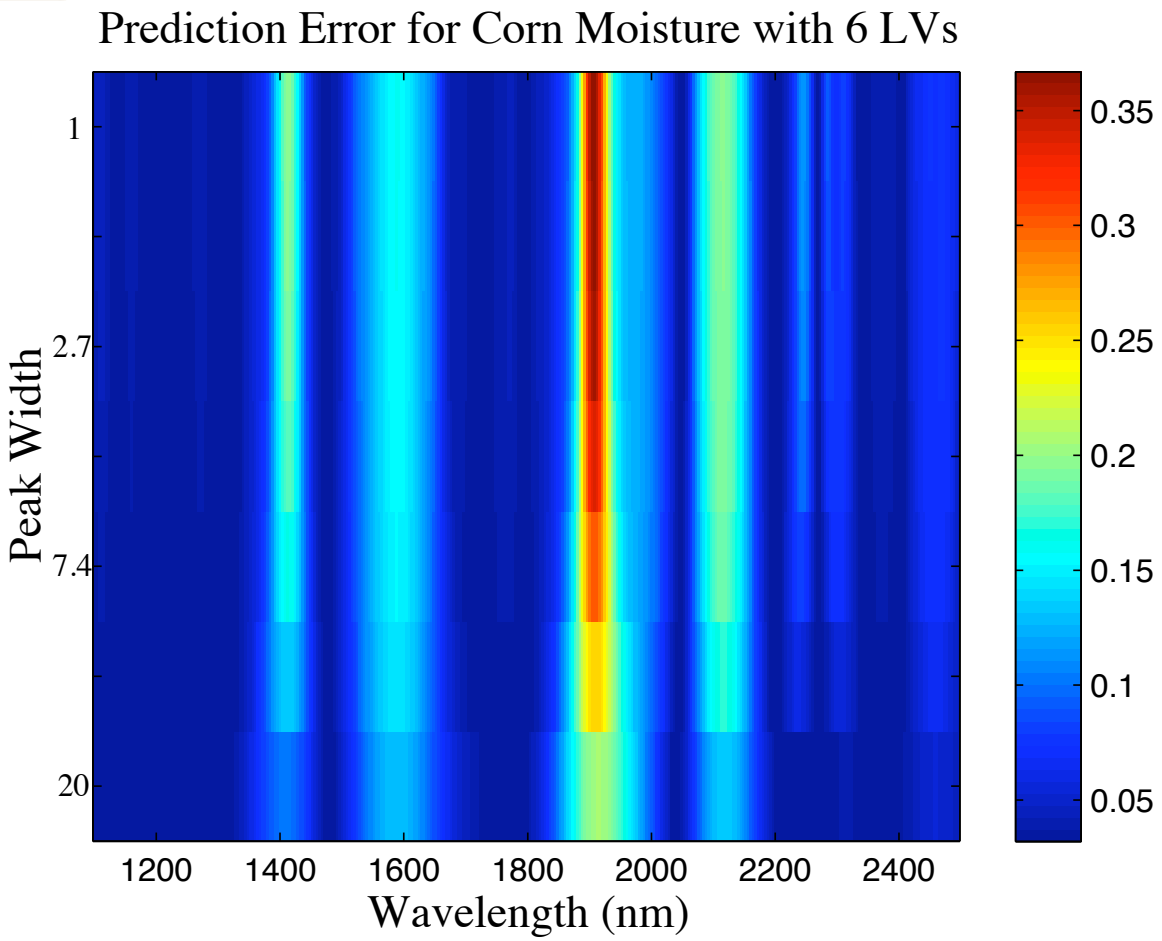
EIGENVECTOR RESEARCH INCORPORATED

Example Data: Corn NIR



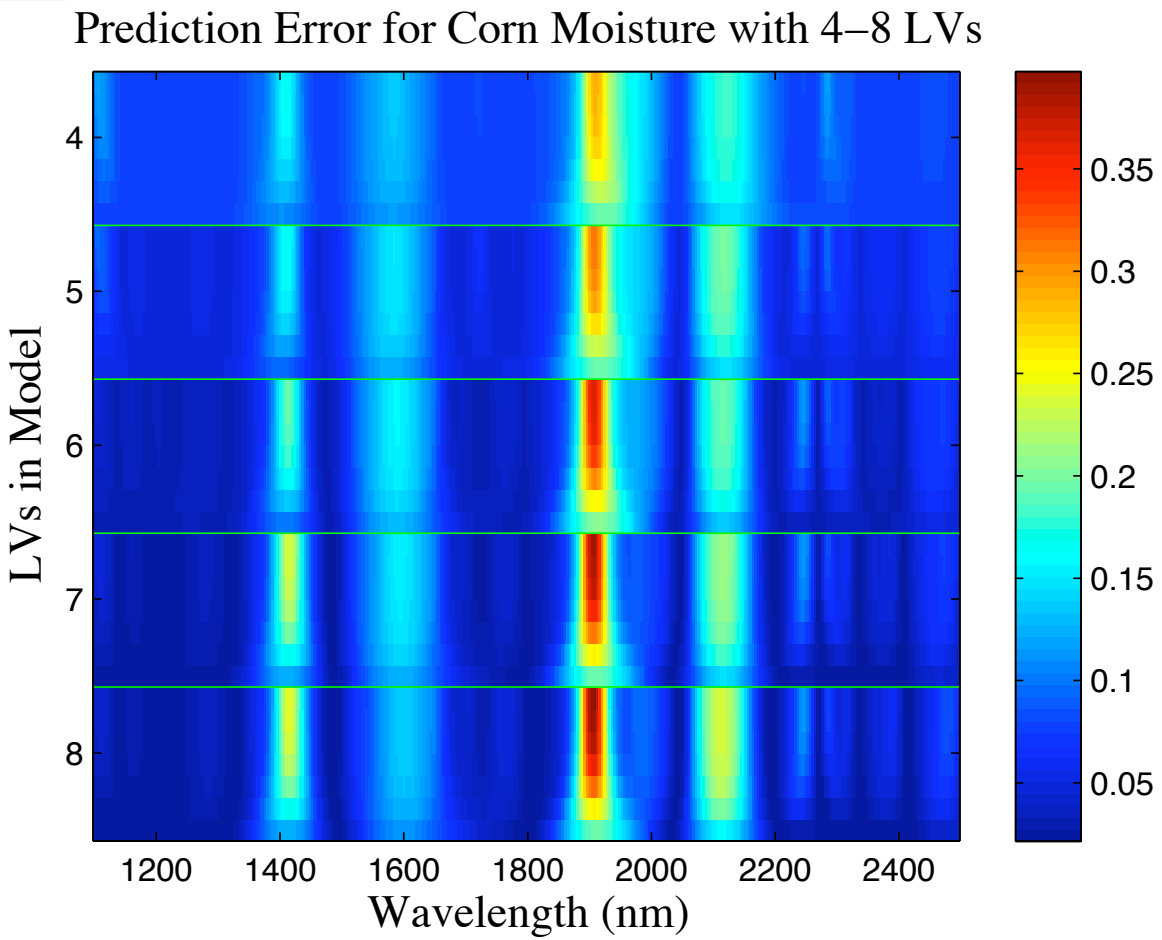
EIGENVECTOR RESEARCH INCORPORATED

Test Corn Model



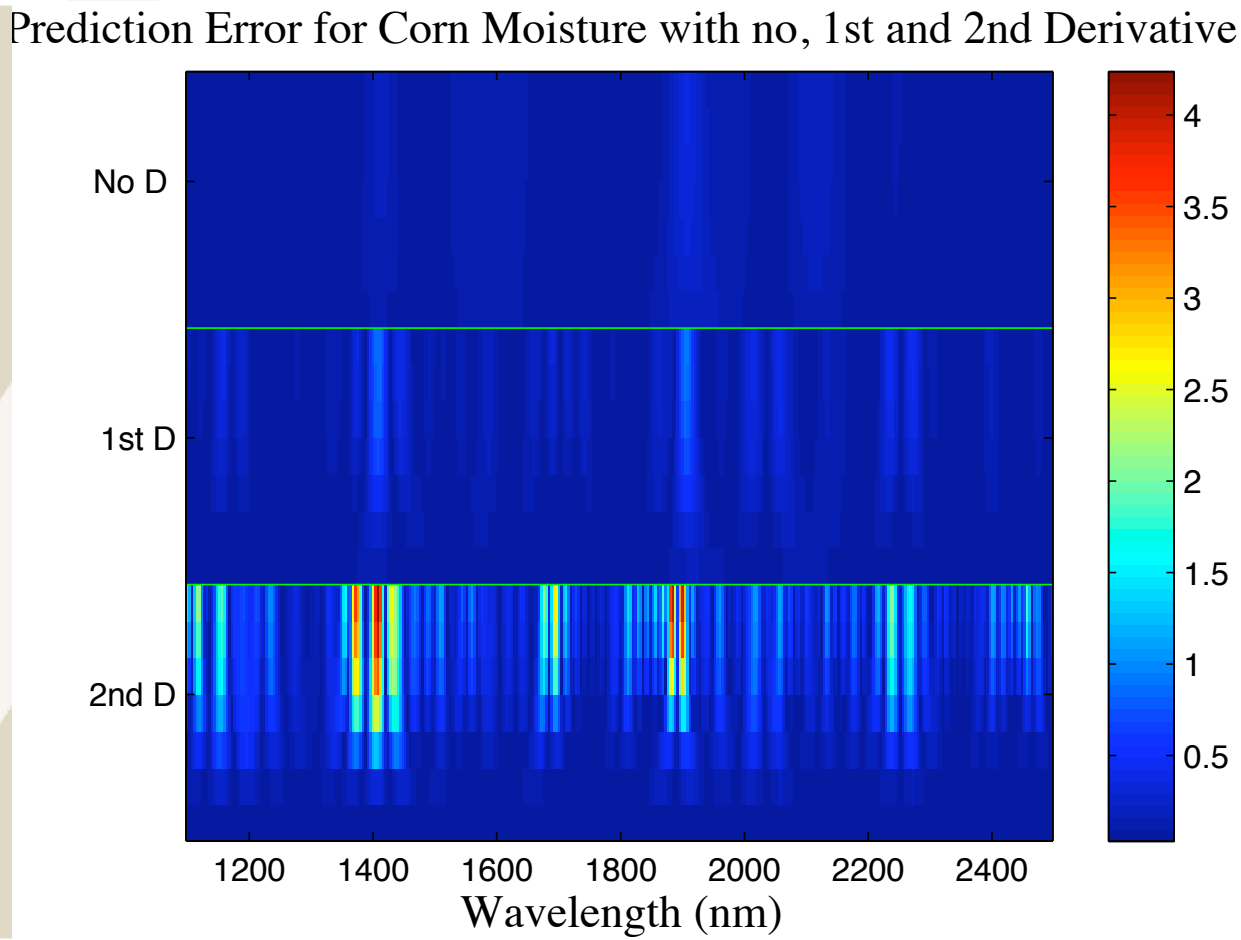
EIGENVECTOR RESEARCH INCORPORATED

Compare Models- #LVs



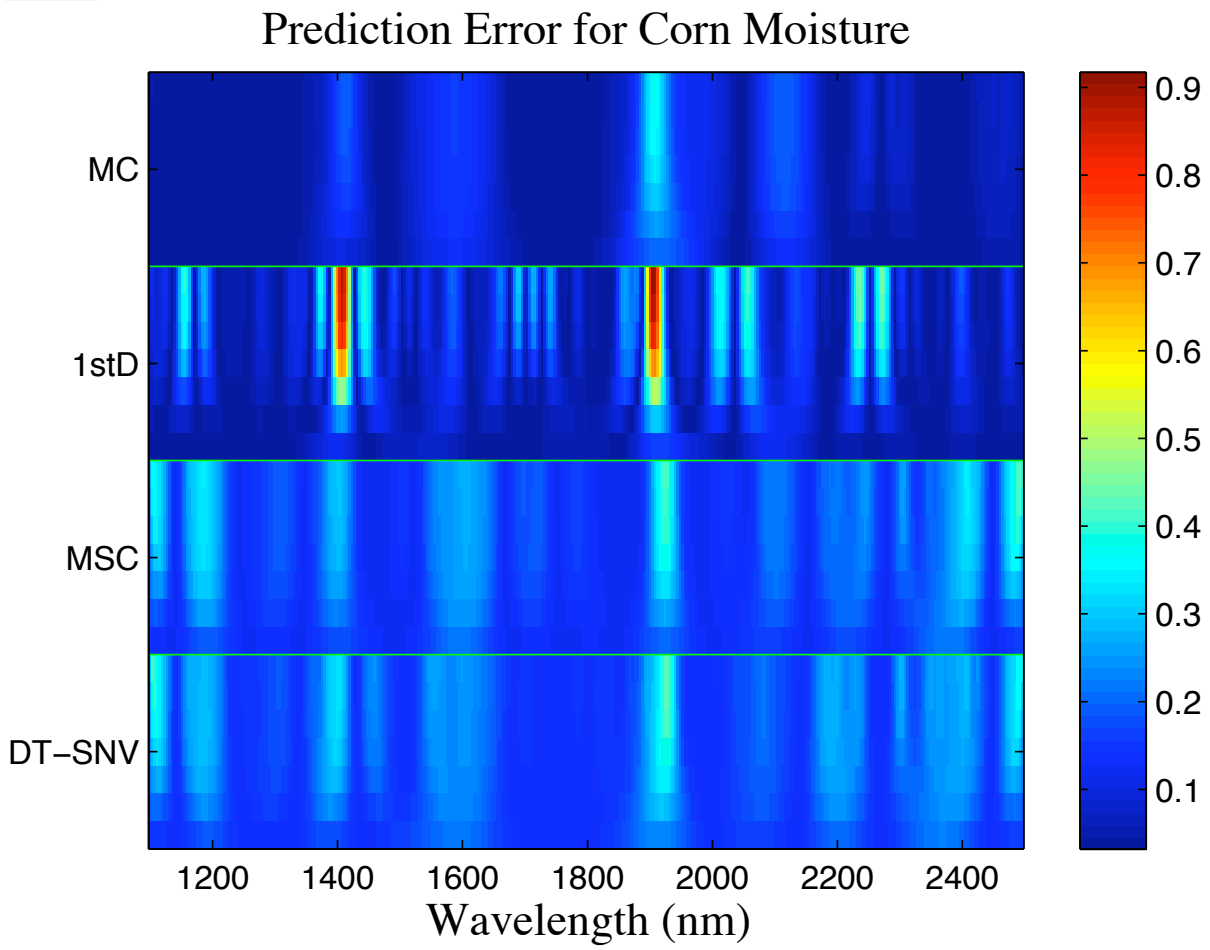
EIGENVECTOR RESEARCH INCORPORATED

Compare Models-Derivative



EIGENVECTOR RESEARCH INCORPORATED

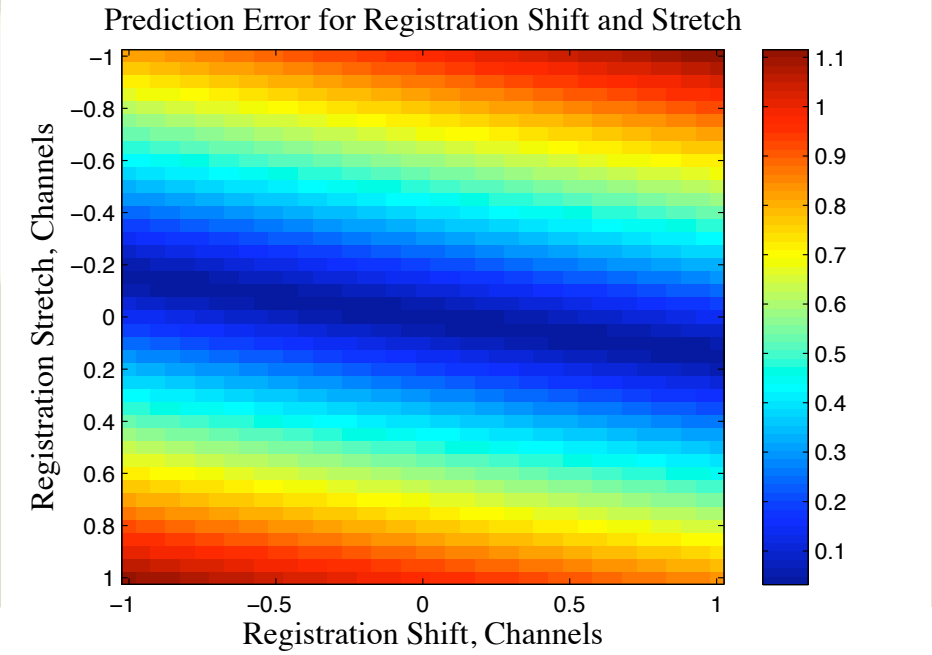
Other Preprocessing



EIGENVECTOR RESEARCH INCORPORATED

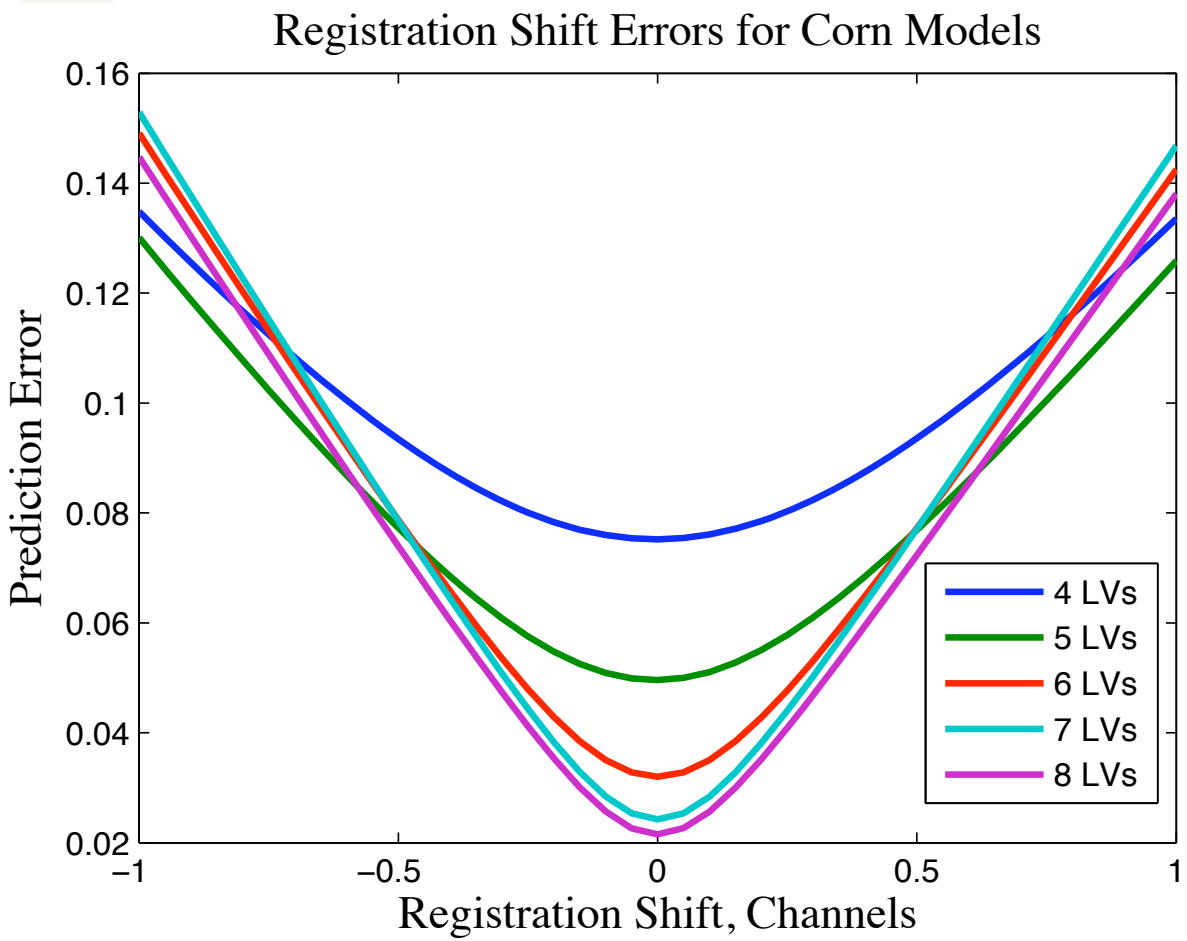
Registration Shift

- Registration shift
  - Left-right
  - Expand and Contract



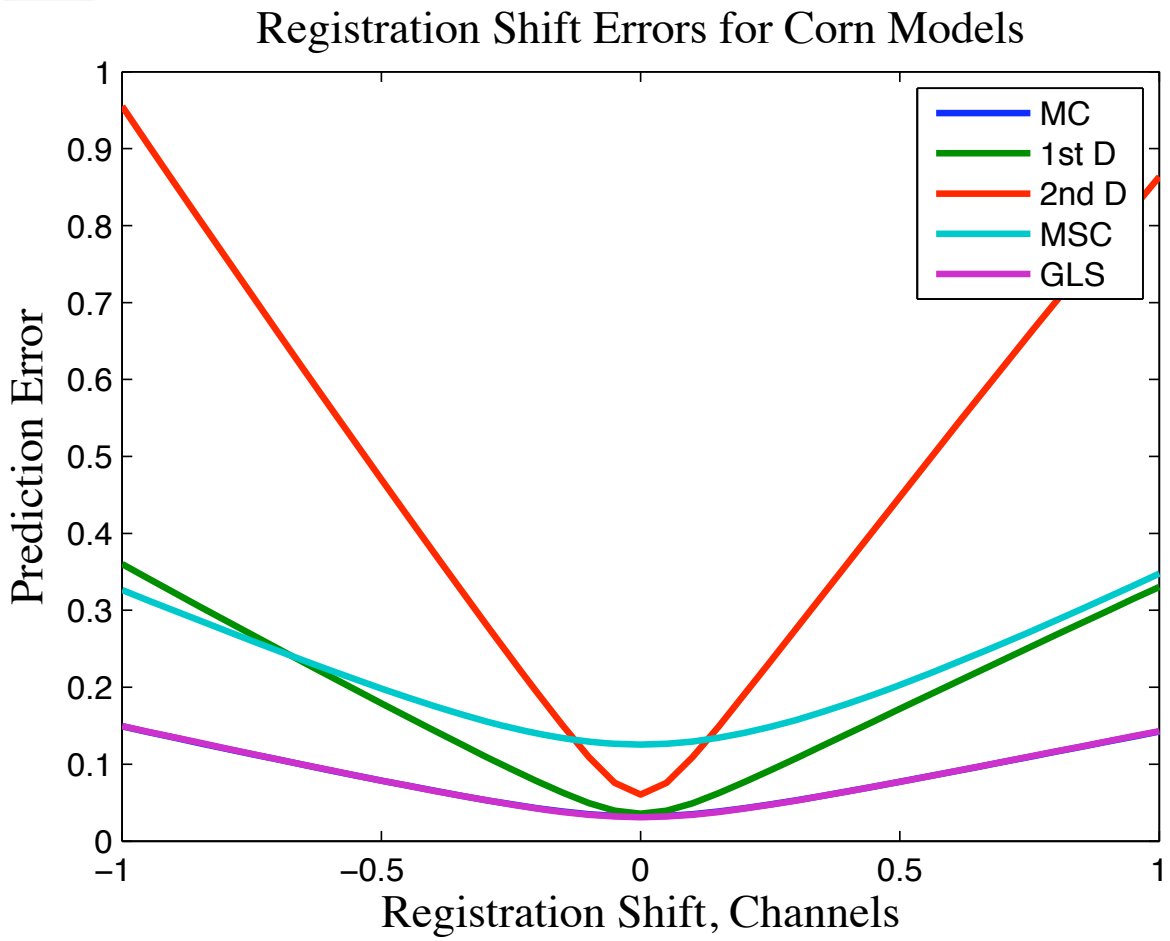
EIGENVECTOR RESEARCH INCORPORATED

Shift with #LVs



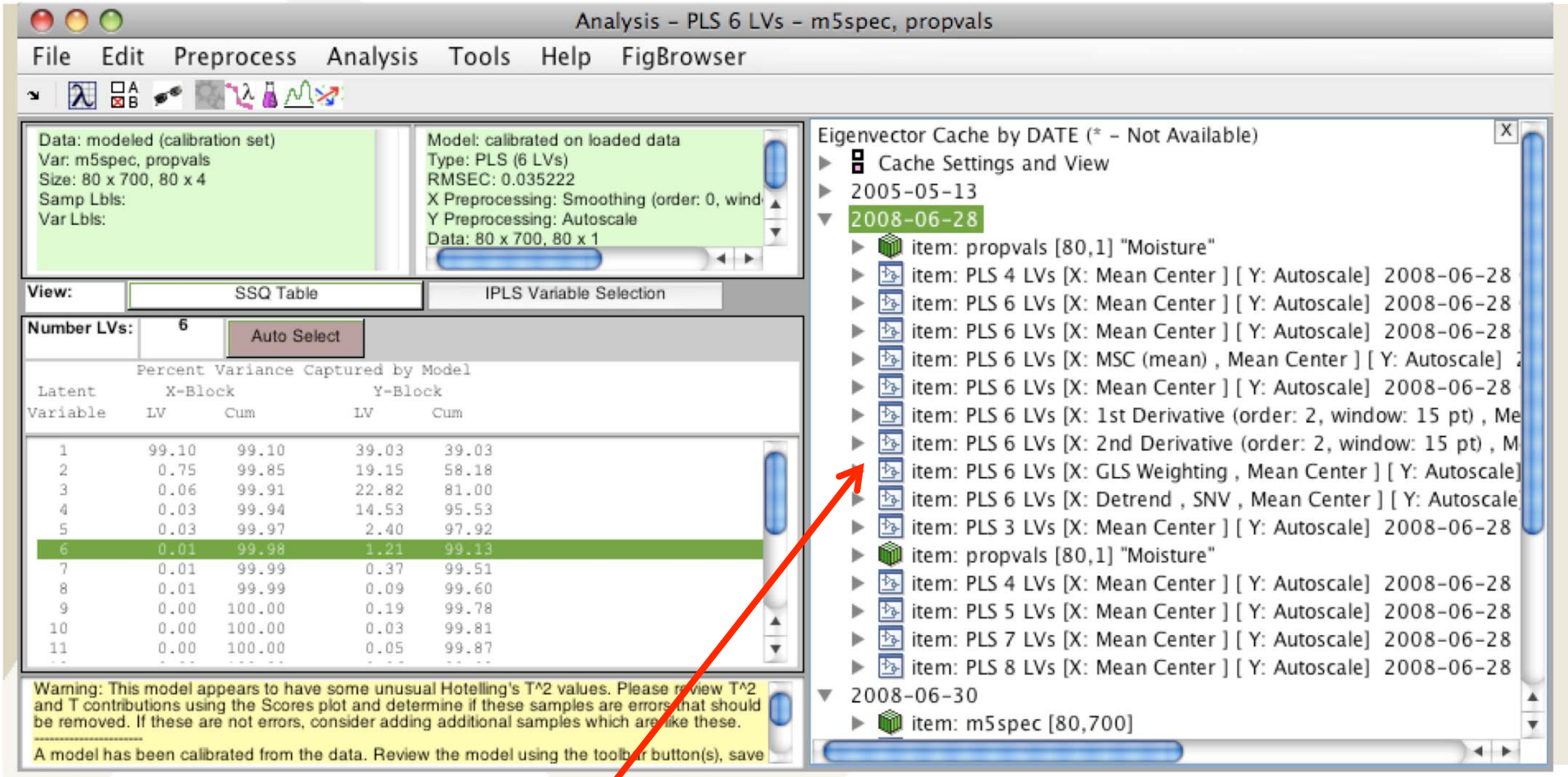
EIGENVECTOR RESEARCH INCORPORATED

Shift with Preprocessing



EIGENVECTOR RESEARCH INCORPORATED

Implementation



Model cache

EIGENVECTOR RESEARCH INCORPORATED

Conclusions

- Functions developed to test model robustness in face of new data "non-idealities"
- Makes it easy to compare models
  - Across #LVs
  - Preprocessing
  - Variable selection
  - Other model types? (Functional?)
- Models more brittle with LVs (knew that)
- Some preprocessing techniques more robust (*e.g.* GLS) than others (*e.g.* 2nd derivative)

EIGENVECTOR RESEARCH INCORPORATED

References

- S.C. Rutan, O.E. de Noord and R.R. Andréa, "Characterization of the Sources of Variation Affecting Near-Infrared Spectroscoy Using Chemometric Methods," *Anal. Chem.* Vol. 70, 3198-3201, 1998.
- F. Estienne, F. Despagne, B. Walczak, O.E. de Noord, and D.L. Massart, "A comparison of multivariate calibration techniques applied to experimental NIR data sets Part III: Robustness against instrumental perturbation conditions," *Chemo. Intell. Lab. Sys.* Vol. 73, 207-218, 2004.

EIGENVECTOR RESEARCH INCORPORATED

Questions?

EIGENVECTOR
RESEARCH INCORPORATED