Using Pre-Whitening Techniques for Process Model Stabilization

Jeremy M. Shaver, Barry M. Wise

Eigenvector Research, Inc.
Seattle/Wenatchee, WA

shaver@eigenvector.com

What is Pre-Whitening?

• Inverse linear regression (e.g. partial least squares) likes random noise
• Non-random systematic interference common and make it work harder for same accuracy.
• More components makes models harder to interpret, introduces more noise
• Goal: Reduce presence of systematic information not relevant to prediction
Pre-Whitening Methods

- Add more PLS components… (brute force)
- Orthogonal Signal Correction (OSC)
- Orthogonal Partial Least Squares (OPLS)*
- Generalized Least Squares Weighting (GLS / GLSW) with y-block sort

* use requires patent license agreement

Orthogonal Signal Correction (OSC) and GLS with y-block Sort

- Use y-block to identify spectral differences which are not apparently related to the “property of interest” (e.g. concentration)
  - use y-block value to sort x-block (spectra)
  - calculate row-wise differences
- Use those differences to create filter and “deweight” those features.
  - y-block differences also calculated to use as weights in covariance calculation
GLS with y-block Sort

Spectra with similar y-block values are near each other and should contain similar features.

Sorting Spectra Using y-block

Sort concentration and use order to sort spectra… Spectra with similar y-block values are near each other and should contain similar features.
Concentration Spectra derivative down columns (both conc. and spec.)

Only differences between spectra with similar y-values are left.
Create covariance from these differences.

Isoflavone in Corn
Steven Wright and James Janni
Pioneer Hi-Bred, Inc., A DuPont Business

- Determine isoflavone content in corn kernels (~ 600-7000 ppm)
- 820 samples from 2000 growing season
- NIR 400-2500 cm$^{-1}$
- Goal: <400 ppm and <10 latent variables?!?

(presented at CAC 2002 conference)
Cross-Validation Results for PLS Models

Cross-Validation Prediction Plots
for [isoflavone] < 5000 ppm
Raman Analysis of Ionic Mixtures from Salt Cake Dissolution

Samuel A. Bryan, Tatiana G. Levitskaia, Serguei I. Sinkov

Pacific Northwest National Labs

- Monitor ionic concentration during dissolution of waste tank salt cakes
- Most ions by classical least squares successful
- OH\(^-\) difficult due to significant overlap with water OH envelope (changes with ionic strength and other effects)

Spectral Changes with [OH\(^-\)]
Before we bother, what’s the advantage?!?

- Does the model respond differently to random noise?
- Does the model respond differently to changes in interferences?
- Is the model differently sensitive to Hotelling’s $T^2$ and/or outliers?

Simulated Process Data

Experiment with different positions and amplitudes of analytical band and different interferences.
High Overlap Simulated Example
(Very Low Signal to Noise Ratio)

Prediction at V. High Noise
(peak to peak SNR = 1)
Cross-Validation at Low Noise
(peak to peak SNR = 5)

Prediction at Low Noise
(peak to peak SNR = 5)
Cross-Validation at High Noise
(peak to peak SNR = 2)

Prediction Error at High Noise
(peak to peak SNR = 2)

Changing Noise and Background
Q Residuals With and Without GLS

SNR = 3

Scores With and Without GLS

SNR = 3
Spectral Changes with [OH⁻]

OH⁻ Cross Validation
Conclusions

- Sensitivity to Outliers and Noise not apparently changed
- GLS can achieve factor reduction in some scenarios in which OSC cannot
- GLS pre-processed data less sensitive to selection of number of factors
- Easier to interpret models may be primary benefit