Determination of Figures of Merit for NIR and Raman Spectrometers by Net Analyte Signal Analysis

Steven M. Short
Robert P. Cogdill, Carl A. Anderson
CAC 2008
July 3, 2008
NIR and Raman Spectra

Calibration
Prevailing methods of Calibration Performance Comparison

- Literature Survey
  - Root-mean-squared error (RMSE)
  - Coefficient of determination ($R^2$)
    - Precision Statistics
    - Signal-to-noise (S/N) ratio

- Are these metrics sufficient or is there additional information that is being ignored?
The **wrong question:**

What SEP do I need for approval (validation)?
The correct question:

What defines performance for this process?

FDA

Pharma

EMEA
Multivariate Net Analyte Signal (NAS) Theory

• Separation of relevant signals from remaining interfering elements

• Avraham Lorber is accredited with its conception (1986)
  ▪ D. R. Morgan published on a similar topic 9 years prior, although the work reportedly contains “some errors”

• NAS provides the platform to compute multivariate figures of merit (FOM)
  ▪ Previously implemented for univariate measures
NAS Theory Continued

• Multivariate NAS first implemented via pure component projection and classical regression
  ▪ Cumbersome in that spectra or concentrations for all contributing species required
• Since been solved in terms of inverse regression
  ▪ Requires only the concentrations for the component(s) of interest
• Mathematically, NAS is the portion of signal related to the constituent being considered
  ▪ Signal directly useful for quantification
  ▪ NAS is orthogonal to all other components within the spectral matrix
Net Analyte Signal Vectors

Analyte Spectrum \((r)\)

Net Analyte Signal \((r^*)\)

Interferences \((r_{\perp})\)

\[ r = r^* + r_{\perp} + \varepsilon \]

Net Analyte Signal Vectors Continued


$\text{NAS} \neq \text{true NAS}$
“Empirical” NAS Computation

\[
\hat{NAS}_i = (x_i \cdot b) \cdot \left( b^T \cdot b \right)^{-1} \cdot b^T
\]

- \( \hat{NAS}_i \) = NAS vector for the \( i^{th} \) sample for a specific component
- \( x_i \) = sample spectrum
- \( b \) = regression vector (column vector)
  - Principal Components Regression (PCR) or Partial Least-Squares (PLS)

NAS Computation Continued

• Net analyte signal can also be expressed in scalar form with no loss in information by

\[
\hat{NAS}_i = \left\| \hat{NAS}_i \right\|
\]

• Where \( \left\| \cdot \right\| \) indicates the Euclidian norm
Figures of Merit (FOM)

- Net analyte signal affords the determination of various figures of merit, providing additional understanding
  - Sensitivity
    - Vector quantity and scalar form
    - Analytical Sensitivity
  - Selectivity
  - S/N ratio
  - Limit of Detection (LOD)
  - Limit of Quantification (LOQ)
Sensitivity

- Characterizes the extent of signal variation as a function of analyte concentration
- Higher sensitivity translates to a greater signal response to a change in concentration

\[
\hat{SEN}_i = \frac{\hat{NAS}_i}{y_i}
\]

- \(\hat{SEN}_i\) = the vector of sensitivities for each instrument variable
- \(\hat{NAS}_i\) = the net analyte signal vector
- \(y_i\) = the measured concentration for the \(i^{th}\) sample
Sensitivity can be expressed in scalar form

\[ \hat{SEN}_i = \left| \hat{SE}_N \right| \]

- Measured in units of instrument intensity per concentration
- Not comparable across all calibrations
Analytical Sensitivity ($\gamma$)

- Developed to provide an impartial assessment between dissimilar analytical techniques

$$\gamma = \frac{\hat{SEN}}{\delta r}$$

$\gamma^{-1}$ = minimum discernible concentration difference given the dynamic range modeled, effective resolution

- $\hat{SEN}$ = mean of sensitivity values
- $\delta r$ = measure (estimate) of instrumental noise
- Measured in units of concentration$^{-1}$

Selectivity

- Dimensionless univariate measure of the portion of signal not lost due to spectral overlap
- Quantity of signal unaffected by the interfering factors

\[ SEL_i = \frac{\| \hat{NAS}_i \|}{\| x_i \|} \]

- Ranges from 0 to 1
- \( \hat{NAS}_i \) = the NAS vector
- \( x_i \) = sample spectrum of the \( i^{th} \) sample
Signal-to-Noise (S/N) Ratio

- Ratio of useable signal to instrumental noise

\[ \frac{S}{N_i} = \frac{\hat{NAS}_i}{\delta r} \]

- \( \hat{NAS}_i \) = scalar NAS value
- \( \delta r \) = measure (estimate) of instrumental noise
- This value has associated units!

\[ S / N_i = \frac{a_1 \cdot (\hat{NAS}_i) + a_o}{\delta r} \]

- \( a_1 \) = scale and \( a_o \) = offset coefficients to transform NAS to units of concentration
**Limit of Detection (LOD)**

- Minimum statistically discernible concentration

\[
LOD = \frac{k_D \sigma}{m}
\]

- \(k_D\) = statistical confidence factor
- \(m\) = slope of the classical least-squares fit of measured versus predicted concentration values
- \(\sigma\) = instrumental noise (\(\delta r\))
- LOD and LOQ differ only in \(k_D\)
Experimental

• Constituents
  ▪ Anhydrous Theophylline
  ▪ Lactose 316 Fast Flo NF Monohydrate
  ▪ Microcrystalline Cellulose (MCC, Avicel PH 200)
  ▪ Soluble Starch

• 29 mixture design points were constructed
Experimental Continued

- 13-mm tablets were produced from each mixture at five compaction levels on a Carver Automatic tablet press
  - Flat-faced punches
  - 10 second dwell time
  - 67.0, 117.3, 167.6, 217.8, 268.1 MPa

- NIR and Raman spectra were collected for all tablets and calibrations were constructed
  - Precision measurements (repeatability and intermediate) were also collected
PLS Calibration Results – Raman

![Graph 1: RMSEC vs Latent Variables](image1)

Square = RMSEC
Triangle = RMSECV

![Graph 2: Predicted Concentration vs Measured Concentration](image2)

Circle = 50th Percentile
Stars = 25th and 75th Percentiles

RMSEC (square), RMSECV (triangle) [%]
Multivariate Selectivity - NIR

Anhydrous Theophylline

Lactose Monohydrate

First Derivative Intensity

Wavelength (nm)

NAS and Interference Spectra

PLS b vector

100% Theophylline

100% Lactose

Center for Pharmaceutical Technology
FOM Results

- NIR and Raman accuracy was highest for theophylline
  - Higher sensitivity and selectivity due to its pure component orthogonality
  - Spectral overlap is manifested in the performance-related FOM (S/N, LOD)
- Dissimilar constituent ordering for highest to lowest analytical sensitivity
  - Pair the instrument to the analytical task
- Sample positioning error had a major impact on the Raman (and NIR) data
  - Spot size, volume interrogated, and sample positioning
- $R^2$ and RMSE are not necessarily predictive of the S/N
  - Accuracy statistics are heavily influenced by the experimental design where as S/N is inherent to the method
- The reduced precision of the Raman data had no negative effect on the ability of PLS to resolve the covariance structure
  - ILS is less affected by the precision than by sample leverage in the estimation of the true solution
Concluding Remarks

• Both technologies provided functional calibrations
• The power of the comparison can be enhanced by determining additional calibration figures of merit
• Calibration Optimization
  ▪ Possible deviation from $R^2$ and RMSE using analytical sensitivity or S/N ratio (future studies)
• Empirical NAS is convenient, but may not meet the needs of the desired task
  ▪ GLS- or PCP-NAS can directly orthogonalize to what is defined as “interference”
Thank you.

Calibration Optimization

![Graph showing Autoscaled Intensity vs. Latent Variables (#)]

- RMSECV
- Analytical Sensitivity