

Development of an NIR Method for Determining the Active Ingredient in Pharmaceutical Tablets

Charles E. Miller

Key words: NIR, pharmaceutical, PAT, PLS, variable selection

Introduction: Multivariate modeling techniques are particularly useful for developing and maintaining process analytical (PAT) methods that involve multivariate analytical instruments [1-3]. In the pharmaceutical industry, one such analytical technology, near-infrared (NIR) spectroscopy, has shown great potential for improving process yields, process safety, environmental compliance and product uniformity.

Contrary to popular perception, the calculation of a predictive calibration model is only one of many different modeling tasks that are required to develop, deploy and maintain effective NIR multivariate calibrations in PAT applications. Some others include data exploration, outlier filtering, variable selection, model interpretation, model testing, and model monitoring. This paper demonstrates the utility of PLS Toolbox for several of these critical tasks.

Experimental: The “NIR Shootout 2002” dataset is used for this study. This data consists of a series of NIR diffuse transmission spectra of intact tablets, collected using a Multitab spectrometer (Foss NIRSystems, Silver Spring MD). Each spectrum covers the spectral range 600 – 1898 nm, in 2 nm increments. Each tablet has an assay value for its active ingredient, which is in the range of 152 to 239 mg, with a nominal value of 200 mg and estimated precision of +/- 1.3 mg. Details regarding this data set can be found in reference [4].

The objective of this analysis is to develop and test an NIR method for determining the assay value of these tablets.

Data Exploration: Simple “exploration” of the raw data, by just plotting the data or building exploratory models on the data, can lead to valuable insight that can be used to guide the user towards an optimal solution. In PLS Toolbox, there are several options for displaying and plotting the raw data, including time-series plots, 2D and 3D scatter plots, variable profile plots, and tabular data display. In this example,

the NIR spectra of the calibration data (Figure 1), and a preliminary PLS model (Figure 2) provide useful information regarding possible outliers, as well as the nature of irrelevant baseline offset and multiplicative variations between the spectra.

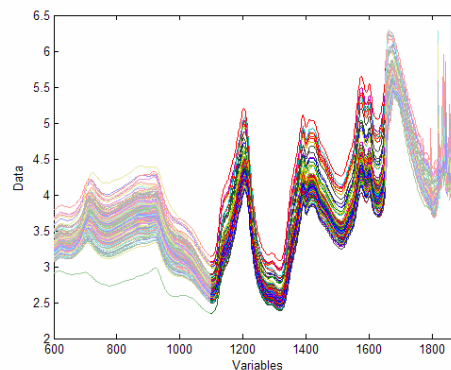


Figure 1: Data Exploration: overlaid profile plot of the NIR spectra in the calibration data set. Grayed-out parts of the plot indicate spectral regions that were excluded from consideration in model development.

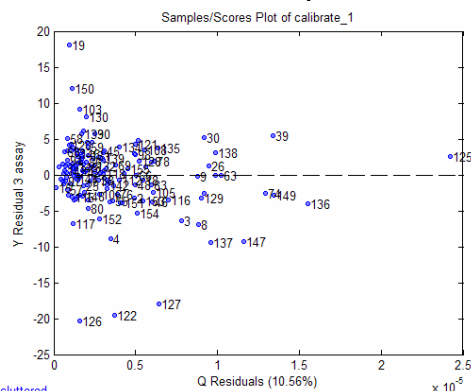


Figure 2: Data Exploration: X- and Y-residuals from a preliminary PLS model on the calibration data-identifying unique samples and possible outliers.

These and other exploratory analyses were used to suggest some wavelength and sample exclusions from the calibration data, as well as some spectral preprocessing strategies.

Data Preprocessing: The highly-flexible preprocessing module in PLS Toolbox allows one to rapidly apply and visually assess a wide range of different options, including “custom” methods. In this example, a specific custom

preprocessing method (2nd derivative, followed by SNV, then mean-centering) is applied.

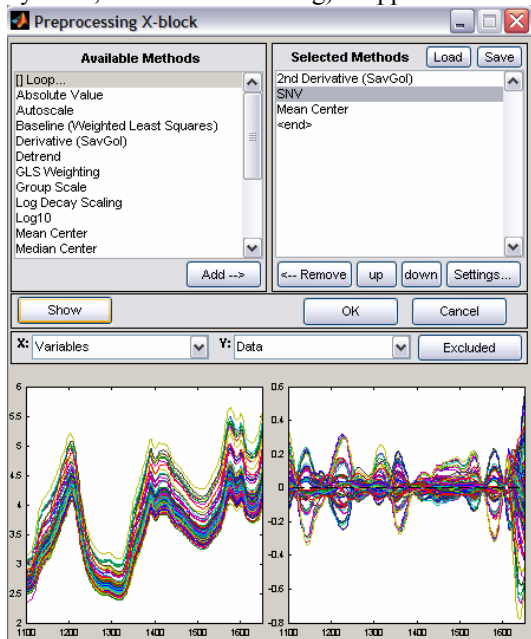


Figure 3: Data Pretreatment: Effect of a “custom” preprocessing (2nd derivative followed by SNV, then mean-centering) on the NIR spectra in the calibration set.

Variable Selection: In NIR, and other spectroscopy applications, variable selection techniques can be used to simplify the calibration model, thus making it less susceptible to unforeseen disturbances and interferences. PLS Toolbox provides several variable selection techniques, including the Genetic Algorithm and iPLS methods [3,5]. Figure 3 shows results obtained from applying the Genetic Algorithm method to the pretreated calibration data.

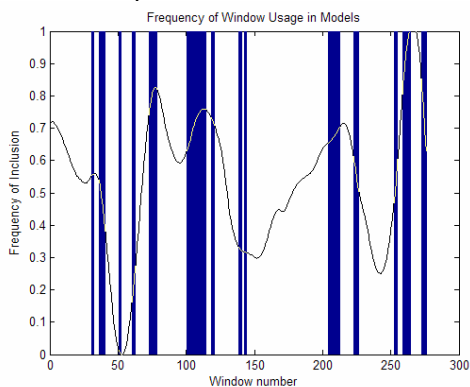


Figure 4: Variable Selection: Results of Genetic Algorithm: selected variables are denoted by blue bars.

Model Building, Interpretation, and Testing:

Once a final model is constructed, it is often advantageous to test it using a set of samples that were not used to build it. In this example, a PLS

model was built using the set of wavelengths shown in Figure 4, and then applied to an independent set of test data. Figures 5 and 6 are a small sampling of the many graphical tools that enable the user to both better understand the final model and to interpret the results of model testing on external data.

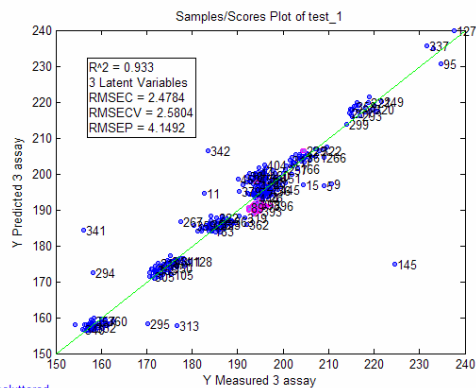


Figure 5: Model Testing: Measured vs. Model-predicted assay values, for external test data. Samples with spectral residual above the 95% limit are selected (pink).

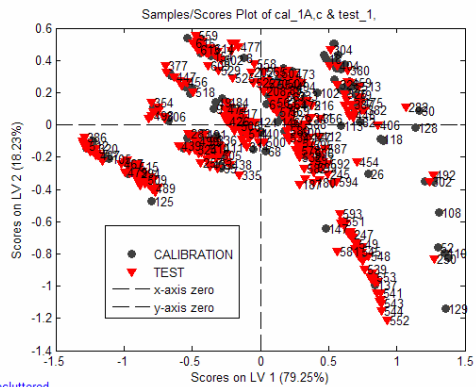


Figure 6: Model Testing: Overlaid scatter plot of PLS scores for LV1 vs. LV2, for both the calibration and test data.

References:

[1] C. E. Miller, Chemom. Intell. Lab. Syst., 30, 1995, 11-22.
 [2] C. E. Miller, J. Chemom., 14, 2000, 513-528.
 [3] C. E. Miller, "Chemometrics in Process Analytical Chemistry", in Process Analytical Chemistry, Blackwell Publishing, Oxford, 2004.
 [4] www.idrc-chambersburg.org/shootout_2002
 [5] R. Leardi, L. Nørgaard, J. Chemom., 18, 486-497, 2004.