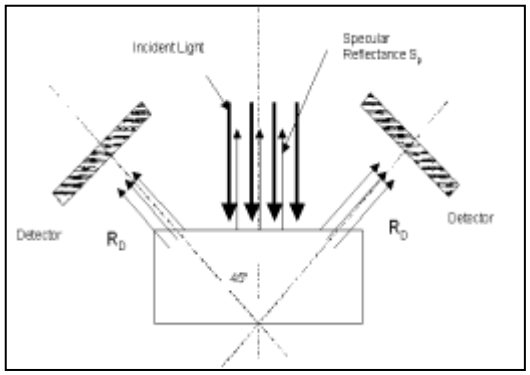
Application Note 180: Discriminant Analysis of Raw Materials in the Pharmaceutical Industry

Introduction:

Near Infrared Reflectance spectroscopy is best performed in the 1900 to 2500nm region of the electromagnetic spectrum. Within this spectral region, Proteins, Amines, Amides (N-H), Moisture, Alcohol, Phenols (O-H) and Aliphatic Hydrocarbons, Aromatic Hydrocarbons Oils, Fats (C-H) absorb NIR energy. Using 0 – 45 degree illumination and detection optics, as shown in figure 1, provides a means of collecting NIR spectra from samples such as the Raw Materials used in Pharmaceutical manufacturing. Using Discriminant Analysis software provides a means to collect diffuse reflectance spectra from powdered samples and make a very accurate and precise means of identifying materials based on their NIR spectrum.



The MultiScan Series 4000 FTNIR Spectrometer offers several advanced features for performing Discriminant Analysis of Raw Materials: including superior wavelength accuracy and precision, variable resolution, faster scanning speed and immunity to stray light or external light.

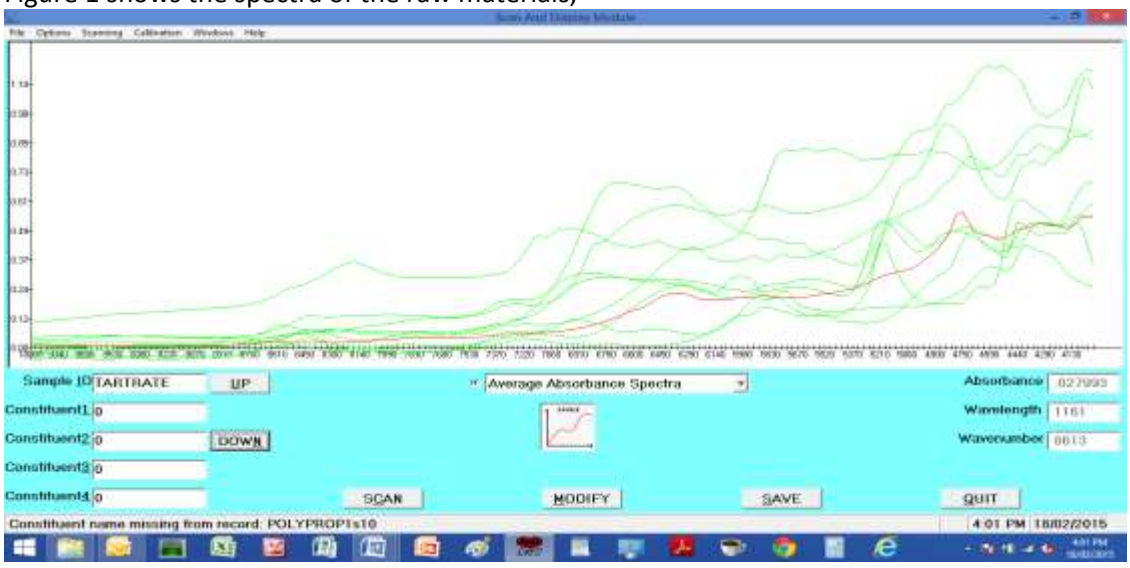
The rotating sample dish provides a means of collecting at least 10 scans from a sample of powdered materials. This ensures a good average spectrum from the sample for comparison with the spectral library of known samples. Typically the Discriminant Analysis takes less than 1 minute.

This application note illustrates the Discriminant Analysis of six materials using a library of materials commonly used in the Pharmaceutical industry.

Procedure:

Samples of raw materials were scanned using the Series 4000 FTNIR Spectrometer and a 10mm deep sample dish. The resolution for the scans was set at 64cm-1. 10 sub scans were collected for each sample and saved as the average spectrum for each sample into separate library files.

Figure 1 shows the spectra of the raw materials;



NTAS (Near Infrared Analysis Software) includes the Discriminant Analysis routine. Figure 2 shows the Discriminant Analysis window.

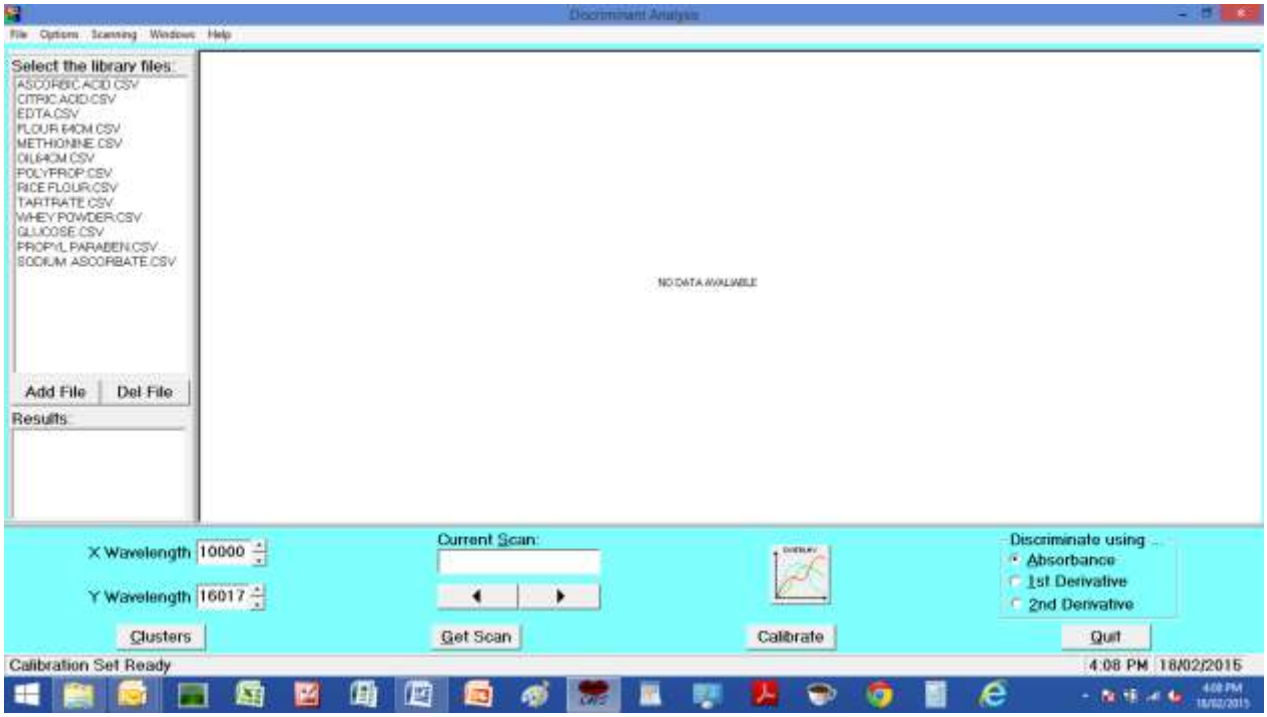


Figure 2. Discriminant Analysis Window in NTAS.

Once the library files have been loaded, the Calibrate button is pressed. The software computes the Covariance Matrix for all the sample files. This is a complex multi dimensional vectorial space that generates a single resultant vector to represent each sample.

When an unknown sample is scanned, the Discriminant Analysis software computes the resultant vector for the new sample scan and then calculates the Mahalanobis Distance from this vector to every other vector in the multi dimensional vectorial space. The lower the Mahalanobis Distance from the unknown sample vector to the other vectors, the more likely that the two vectors are from the same sample.

The software then shows a list of 5 samples with an increasing Mahalanobis Distance between the library spectrum and the unknown spectrum. A perfect match should see a Mahalanobis Distance of 0, however the number is always a positive number.

Results:

Figure 3 through 9 show the results of performing the Discriminant Analysis routine on seven samples of materials which are included amongst the library files, ie, EDTA, Glucose, Sodium Ascorbate, Methionone, Poly Parabenzaldehyde and Whey Powder.

The Mahalanobis Distance for each analysis are listed below.

Selected Material	Mahalanobis Distance	Next Material	Mahalanobis Distance
EDTA	7.38	Poly Parabenzaldhyde	114479
Poly Parabenzaldhyde	6.99	Methionone	2296717
Glucose	26.34	Poly Parabenzaldhyde	2249331
Whey Powder	15.80	Poly Parabenzaldhyde	1170226
Methionone	6.03	Sodium Ascorbate	1465663
Citric Acid	19.51	Poly Propylene	707594

Figure 3. EDTA

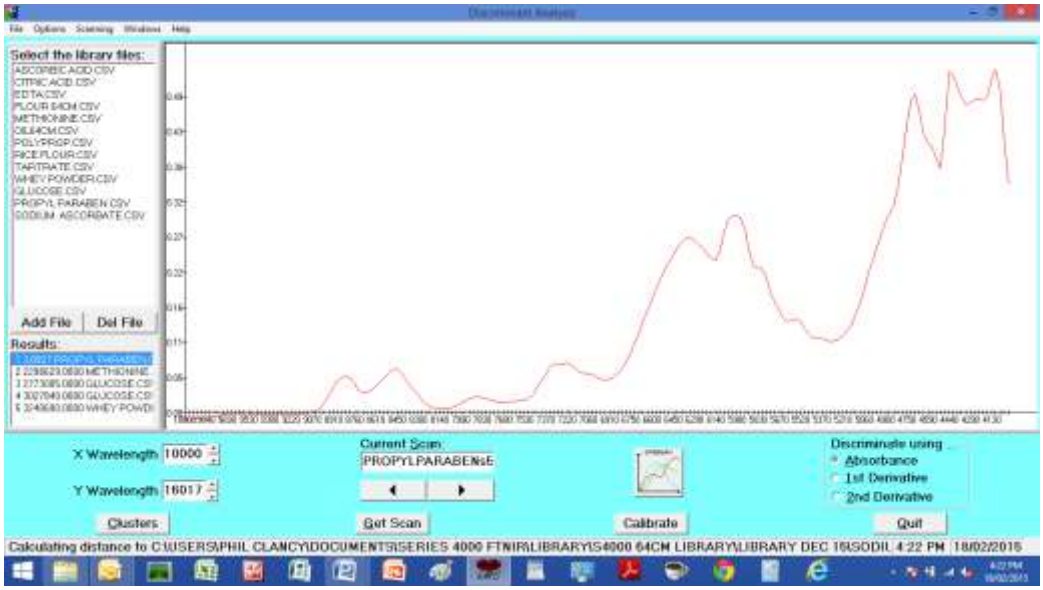


Figure 4. Poly Parabenzaldehyde

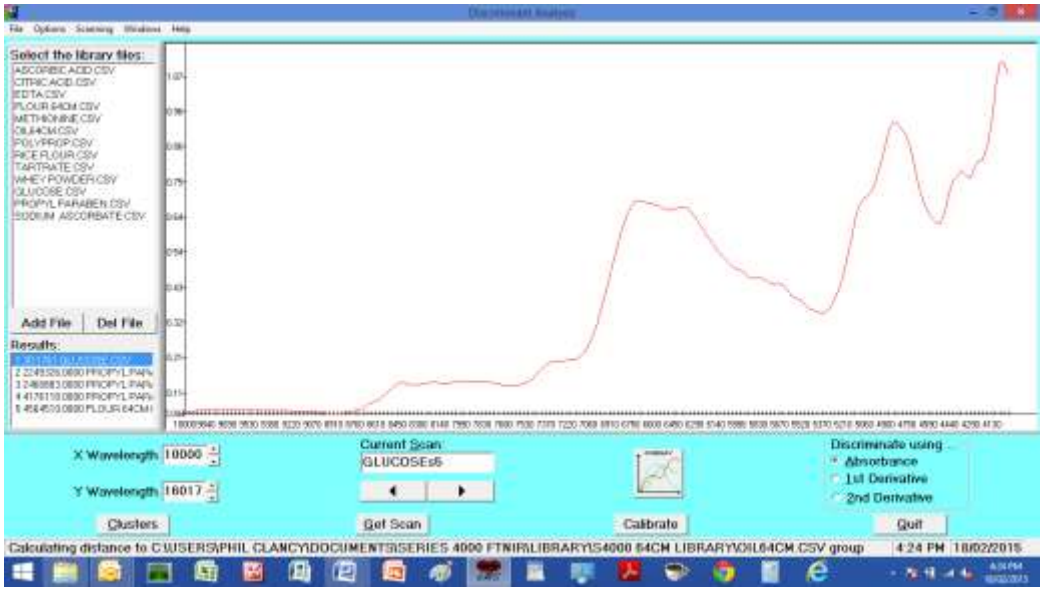


Figure 5. Glucose

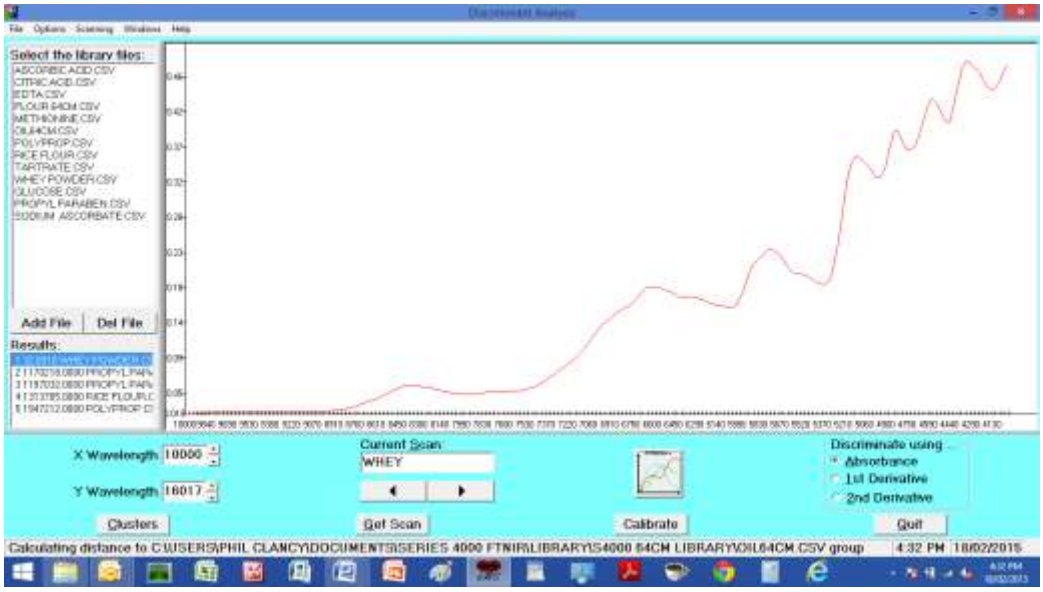


Figure 6. Whey Powder

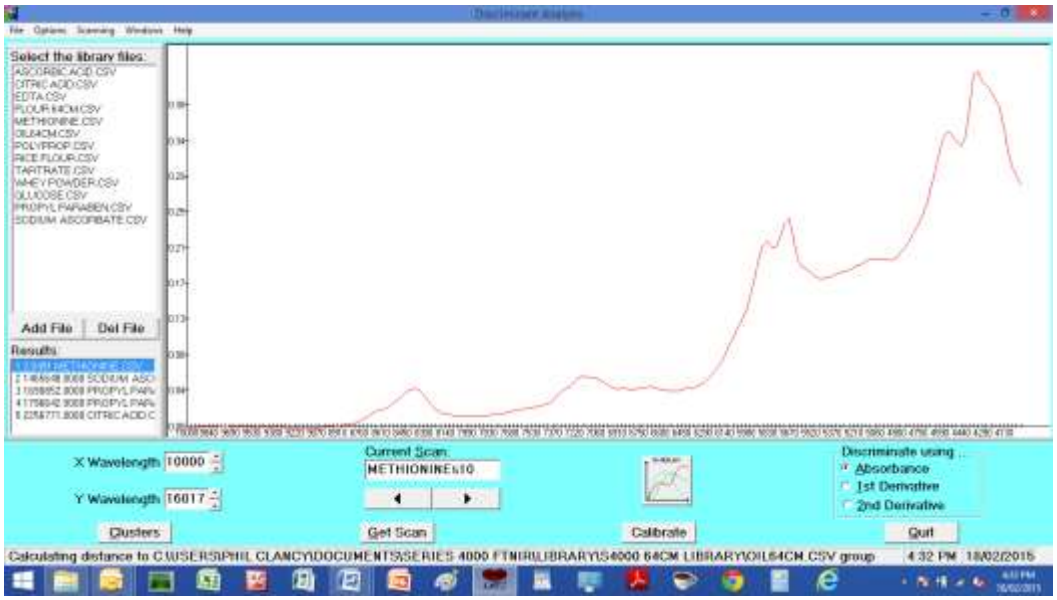


Figure 7. Methionone

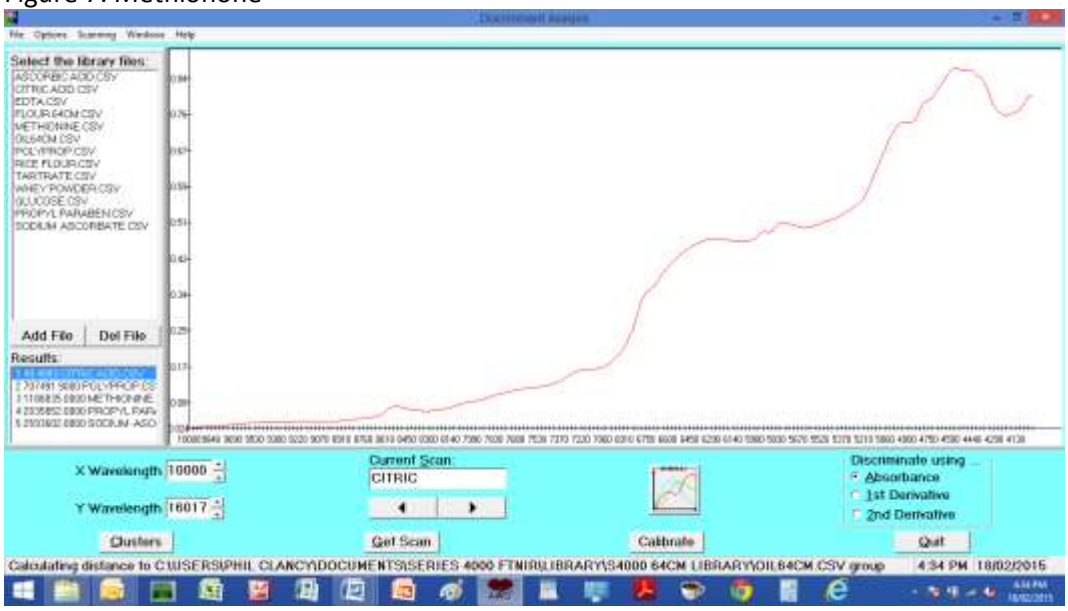


Figure 8. Citric Acid

Discussion:

Discriminant Analysis is a powerful tool to perform library searches and to identify materials based on their Near Infrared spectra. The process of developing a library file is very simple. Scan 3-5 examples or batches of a material so that the spectra represent the day to day and batch to batch variations in the spectral scans. Store these into a unique library files. Over a period of time, add more spectra from different batches and different suppliers in order to build up a robust library of sample spectra. However a set of 30-100 library files can be developed in a few days using retained samples from batches of each raw material.

To analyse a sample using the Discriminant Analysis routine is as simple as filling the sample dish and pressing the Scan button. The software will collect the scans and display the results within 60 seconds.

The results shown in this study illustrate the very large differences in Mahalanobis Distances between different raw materials. It is therefore considered that variations in batches will not result in a false identification.

The Discriminant Analysis software can also be used to define a range of Mahalanobis Distances within which a specific raw material is considered Accept or Reject. For example, if the identification is correct but the Mahalanobis Distance is greater than a set value, ie, 2000, then the software can show a Red Button with Reject. If the Mahalanobis Distance is less than the set value, the software displays and Green Button and Accept.

NIR Technology Systems
366 Edgar Street, Condell Park, NSW, 2200, Australia
Tel: 612 9771 5444 Fax: 612 9771 5255
Email: sales@nextinstruments.net Web: www.nextinstruments.net