APPLICATION NOTE

FT-NIR Spectroscopy

A Comparison of Fiber Optic and NIR Sampling for SIMCA Classification of Spectroscopically Similar Pharmaceutical Raw Materials

Summary

This note compares two different sampling approaches used in FT-NIR spectroscopy to discriminate between seven similar raw materials used in the pharmaceutical industry. These materials were first sampled using a hand held diffuse reflectance fiber optic probe, and were readily discriminated using a SIMCA model. The procedure was repeated using the fixed, fiberless Spectrum™ One Near Infrared Testing System (NTS) fitted with a NIRA accessory, which was found to provide superior data to that collected using fiber optics.* The procedure was repeated using the NIRA reflectance accessory replacing the fiber optic sampling.

Introduction

NIR Spectroscopy is a useful measurement throughout various stages of the manufacturing process, but is particularly useful for raw materials checking and verification. If the materials to be identified are spectroscopically dissimilar, it is often only necessary to use a simple distance measure such as a spectral difference. If the spectra are similar, it may be necessary to use more sophisticated techniques which take into consideration both the variability of the spectra and the differences between the spectra. The SIMCA (Soft Independent Modeling by Class Analogy) technique provides such an example.

*Spectrum One and NIRA have been superceded by the Frontier FT-NIR and NIRA systems.

PerkinElmer®
For the Better

Experimental

Seven different samples were supplied; three different grades of Eudragit®, and four different types of Povidone powders. The NIR spectra of all the samples were recorded on a Spectrum One FT-NIR spectrometer* fitted with an internal fiberless reflectance accessary (NIRA) and a remote diffuse reflectance fiber optic probe. The analysis was performed initially with the fiber optic probe, and then repeated using the NIRA to compare their discriminatory abilities.

Twelve spectra per product were recorded from the batches provided. Products which had two batches contained six replicate spectra, whereas those with three samples contained four replicates. When using the NIRA, the replicates were generated by shaking the vial before re-scanning on the sample platform.

Details regarding the sampling and the scan conditions are given in Table 1. For each of the seven products, seven QUANT+™ methods were built and calibrated. SIMCA diagnostics and validation tests were performed to check the validity of the model. SIMCA modelling is now available using the AssureID™ software.

Table 1. **Summary of the scanning conditions and accessories used.**

Range	12000 – 3800 cm^{-1}
Resolution	16 cm^{-1}
Scanning time	< 1 minute
Accessory	NIRA or fiber optic probe

Results and Discussion

Representative spectra of the products recorded on the NIRA are illustrated in Figure 1. The curves have been offset for clarity, however, it is clear that some of the spectra are visually very similar.

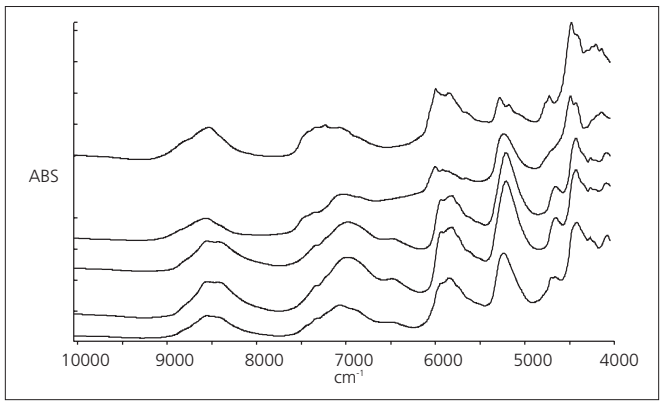

Figure 1. Typical spectra representing different materials recorded on the NIRA.

A global principal components analysis (PCA) was performed on the two data sets collected by different sampling techniques (84 spectra each). The data was found clustered.

Figures 2a and 2b show the data recorded using the fiber optic probe with no pre-processing. After baseline correction was employed (Figure 2b) each of the groups can clearly be seen.

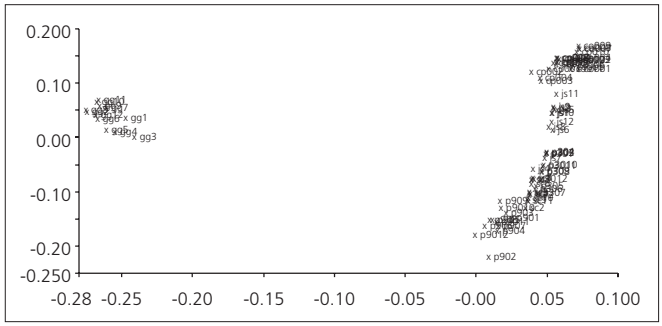

Figure 2a. First two PC scores for 84 test spectra recorded with the fiber optic probe with no pre-processing.

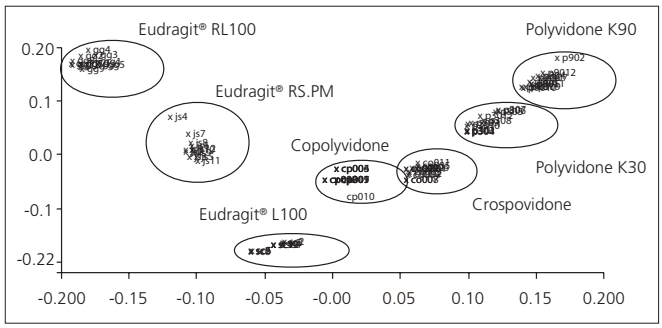

Figure 2b. First two PC scores for 84 test spectra recorded with the fiber optic probe with spectrum pre-processing.

Figures 3a and 3b show the data recorded on the NIRA. Here the data clusters with no pre-processing parameters. Figure 3b clearly shows separation improvement after the pre-processing parameters were added; especially between the povidone groups.

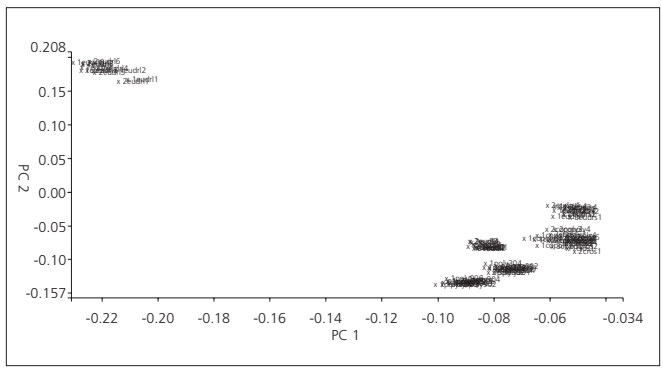

Figure 3a. First two PC scores for 84 test samples recorded on the NIRA with no pre-processing.

** The Spectrum One FT-NIR is superceded by the PerkinElmer Frontier FT-NIR system.*

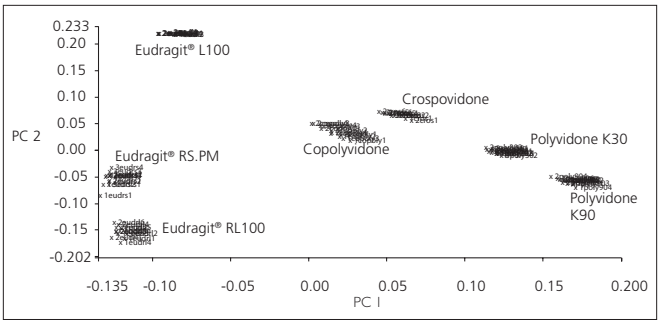

Figure 3b. First two PC scores for 84 test samples recorded on the NIRA with spectrum pre-processing.

Good separation was found using a 9 point first derivative. Since two factors were used in the global PCA it was decided that twelve spectra per product was adequate to use per product. For validation purposes the data was then split into two libraries – one set to be used to build the model and the other set to be used as an independent validation set to test the model. A sample from each batch of product was removed from the data set and used for testing purposes.

Diagnostics

SIMCA methods were built and calibrated for each of the products (7) using the data collected from the fiber optic probe. This procedure was repeated using the data collected from the NIRA. The data set was evaluated using the SIMCA Diagnostics to ensure that none of the batches of spectra overlapped. The diagnostics report provides the interclass distances, i.e. the arbitrary distances between each of the classes.

The procedure also checks every standard spectrum to ensure that the ones from a single class fit that class (recognition), and that those from other classes selected are rejected (rejection). The two rate columns should ideally report 100% for each instance. For both data sets the two rate columns reported 100%, indicating good separation of each class of compound.

An example report is shown in Tables 2a and 2b and clearly show that better separation of the classes can be achieved when using the NIRA sampling technique.

***Table 2a.* Verification Diagnostic Report – probe data set.**

Critical probability level: 0.01 – Inter class Distances

	PolyK30	PolyK90	EudL	EudRL	EudRSPM	Copol	Cros
PolyK30	–	51.12	8475.5	2984.23	5925.77	3110.07	51.95
PolyK90	–	–	9164.6	3309.71	5738.38	4247.97	67.65
EudL	–	–	–	2102.98	3936.68	4670.86	5325.98
EudRL	–	–	–	–	185.08	7173.7	5117.5
EudRSPM	–	–	–	–	–	4834.05	6792.94
Copol	–	–	–	–	–	–	1762.14
		% Recognition rate		**% Rejection rate**			
PolyK30		100(9/9)		100(57/57)			
PolyK90		100(10/10)		100(56/56)			
EudL		100(9/9)		100(57/57)			
EudRL		100(10/10)		100(56/56)			
EudRSPM		100(9/9)		100(57/57)			
Copol		100(9/9)		100(57/57)			
Cros		100(10/10)		100(56/56)			

Table 2b. Verification Diagnostic Report – ICRA data set.

Critical probability level: 0.01 – Inter class Distances

	EudRSPM	Cros	EudL	PolyK90	PolyK30	Copol	EudRL
EudRSPM	–	20418.02	4979.92	11441.76	9658.11	5184.68	227.28
Cros	–	–	15675.03	451.74	229.05	4507.89	17588.41
EudL	–	–	–	17569.45	14735.8	9377.86	3830.74
PolyK90	–	–	–	–	93.76	3348.99	8421.63
PolyK30	–	–	–	–	–	2054.48	8699.67
Copol	–	–	–	–	–	–	3801.37

Key: *Cros = Crospovidone • Copol = Copolyvidone • EudL = Eudragit® L100 • EudRL = Eudragit® RL100*
EudRSPM = Eudragit® RS.PM • PolyK30 = Polyvidone K30 • Polyk90 = Polyvidone K90

Validation

Following the results from the diagnostic procedure, the two data sets were tested using the validation procedure in the SIMCA analysis. This analysis validates the methods that have been built using different test spectra. The models for this case were tested using the spectra that were removed from the data; i.e. the independent validation set. This procedure classifies the spectra and reports the number of misclassifications.

Classification

Very promising results were obtained from the diagnostic and validation procedures of the SIMCA model with both the data collected from the fiber optic probe and the NIRA. The final step was to test unknown spectra against each class.

A spectrum of Eudragit® RS.PM from the independent validation was tested against the probe data set. A corresponding spectrum of Eudragit® RS.PM was tested against the NIRA data set. The results are given in Tables 3a and 3b. The report produced values for the spectrum residuals which are a measure of the lack of fit of the spectrum to the class model. The smaller the number, the more likely the spectrum belongs to that class.

In a similar manner, a number is generated for the model residual which represents the residual within class space. These two figures are combined by the root of the sum of the squares of the spectrum and model residuals as shown in the report. The critical probability level was set to 0.01. Therefore any number produced in the probability column larger than 0.01 is a positive classification, and the 'unknown' belongs to that class.

Table 3a. **Verification Classify Report – Probe**

Description: Eudragit® RS.PM (12549) 6A00970

Class Identification: EudRSPM.MD

Title: Eudragit® RS.PM

Critical Probability level: 0.01

Distance to class	Residual	Model	Combined	Probability
Class Name				
EudRSPM.md	1.332	0	1.332	0.1181
EudRL.md	6.339	3.054	7.037	0
PolyK90.md	63.55	7.698	64.02	0
EudL.md	71.27	2.377	71.31	0
Copol.md	74.24	7.569	74.63	0
PolyK30.md	76.96	7.824	77.36	0
Cros.md	99.53	12.46	100.3	0

Table 3b. **Verification Classify Report – NIRA**

Description: Eudragit® RS.PM 6A00970

Class Identification: EudRSPM.MD

Title: Eudragit® RS.PM NIRA

Critical Probability level: 0.01

Distance to class	Residual	Model	Combined	Probability
Class name				
EudRSPM.md	0.5663	0	0.5663	0.9738
EudRL.md	11.59	7.223	13.66	0
Copol.md	78.62	7.845	79.01	0
EudL.md	88.45	19.31	90.53	0
PolyK90.md	103.4	37.22	109.9	0
PolyK30.md	103.3	51.59	115.5	0
Cros.md	213.8	15.9	214.4	0

(Table 3a) For the data collected via the probe, the probability of the spectrum belonging to the Eudragit® RS.PM class is 0.1181, much bigger than the 0.01 limit. All other probabilities were zero. Therefore, it can be concluded that this model has positively classified the 'unknown' spectrum to be Eudragit® RS.PM.

However, Table 3b presents the same results from the data collected on the corresponding spectrum recorded with the NIRA. Here, the probability is much higher at 0.9738. Therefore, it represents much better classification for the samples recorded on the NIRA rather than the samples recorded using fiber optics.

These results are also presented graphically, as shown in Figures 4a and 4b. The unknown spectrum is represented by the origin, and its distance from each product in the library compared with the critical probability distance. The closer the product is to the origin, the more likely the unknown belongs to that method.

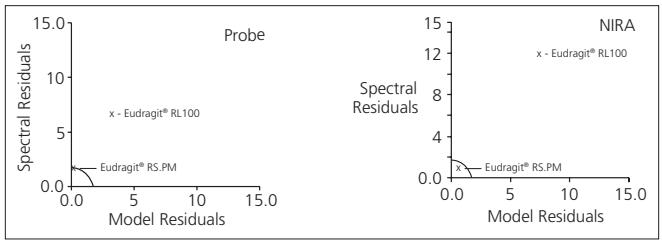

Figure 4a.

Figure 4b.

Table 4a. **Verification Classify Report – Probe**				
Description: Polyvidone (12644) 6A01506				
Class Identification: PolyK30.MD				
Title: Polyvidone K30				
Critical Probability level: 0.01				
Distance to class	**Residual**	**Model**	**Combined**	**Probability**
Class name				
PolyK30.MD	1.066	0	1.066	0.3598
PolyK90.MD	6.945	2.797	7.487	0
Cros.MD	8.088	5.718	9.905	0
EudRL.MD	24.54	4.475	24.95	0
Copol.MD	60.19	1.361	60.2	0
EudRSPM.MD	69.84	23.81	73.79	0
EudL.MD	107.7	16.28	108.9	0

Table 4b. **Verification Classify Report – NIRA**				
Description: Polyvidone K30 6A01506				
Class Identification: PolyK30.MD				
Title: Polyvidone K30 Method				
Critical Probability level: 0.01				
Distance to class	**Residual**	**Model**	**Combined**	**Probability**
Class name				
PolyK30.md	0.7974	0	0.7974	0.7794
PolyK90.md	7.59	8.443	11.35	0
Cros.md	20.87	9.311	22.86	0
Copol.md	57.48	20.31	60.96	0
EudRL.md	70.06	4.01	70.17	0
EudRSPM.md	105.1	49.99	116.3	0
EudL.md	157.7	97.91	185.6	0

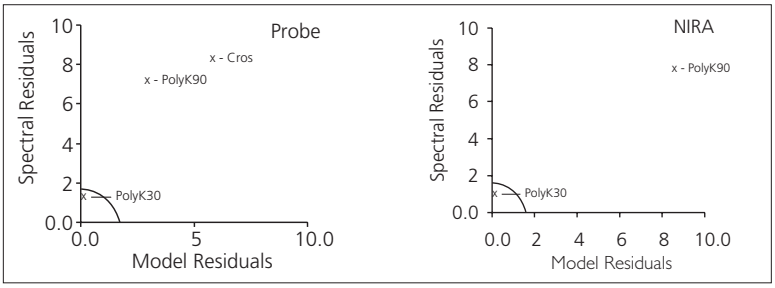

Figure 5a.

Figure 5b.

Therefore, using the examples mentioned above, the unknown is most likely to belong to the Eudragit® RS.PM or RL100 classes. If the class is situated within the quadrant defined by the arc, then the unknown belongs to that class. In this Figure there is conclusive evidence that the unknown belongs to the Eudragit® RS.PM class. However, this evidence is much stronger looking at the NIRA data set due to the much larger separation between the two methods.

Similarly, a second classification test was performed in an attempt to discriminate samples of Polyvidone K30 from the rest of the set. A sample of Polyvidone K30 was taken from the independent validation set and challenged in the model. The classification report is shown in Table 4a and 4b and again different results are produced depending upon the sampling type. A probability of 0.3598 was obtained with the data collected from the probe. This was improved to 0.7794 when the corresponding spectrum collected on the NIRA was tested. This proves that the NIRA is a more reliable and reproducible method of sampling. Again these results are represented graphically and are illustrated in Figures 5a and b.

Conclusion

The SIMCA method is a powerful tool for classifying pharmaceutical products which are spectroscopically similar. This example shows successful classification of materials which differ largely in their physical properties. The NIRA sampling accessory provides a rapid and convenient means of sampling which is free from some of the limitations of fiber optic sampling and this is shown by the improved discrimination using SIMCA when compared with the fiber optic probe.

PerkinElmer, Inc.
940 Winter Street
Waltham, MA 02451 USA
P: (800) 762-4000 or
(+1) 203-925-4602
www.perkinelmer.com

PerkinElmer®

005628B_01