Reduced Variable Multivariate Analysis for Material Identification with the NanoRam®-1064

Raman spectroscopy is a widely used technique for rapid material identification and verification based on the chemical signature that is the Raman spectrum. The advent of handheld Raman instrumentation, often operated via an intuitive touch screen interface, over a decade ago has aided pharmaceutical manufacturing companies in meeting requirements for 100% incoming material inspection. Raman identification testing can be done rapidly and nondestructively, with measurements made through transparent packaging such as plastic drum liners, plastic sample bags, and glass vials. Raman spectroscopy has been shown to be a very effective means of material identification. With the introduction of the NanoRam-1064 that uses a 1064-nm laser excitation, darker materials and those that exhibit strong fluorescence at shorter excitation wavelengths can be effectively identified. Here we describe a new reduced variable multivariate (RVM) algorithm implemented on the NanoRam-1064 that gives it robust methodologies for rapid, nondestructive sample verification.

Handheld Raman spectroscopy provides a near immediate pass/fail or exact identity result, as the collected Raman spectrum is compared either to a specific method or to a library of a variety of different samples, respectively. The user views a clear, unambiguous answer, with the option to view the collected Raman spectrum that forms the basis of the result. There are two common approaches for material identification: a correlation-based method comparing a measured spectrum to embedded library spectra to get a match based on the highest hit quality index (HQI), or multivariate methods based on classification and principal component analysis (PCA) with defined thresholds (typically a statistical p-value of ≥0.05) for a material to be accepted as a member of a class based on a pass/fail result.[1-6]

Raman spectra are multivariate, containing hundreds to thousands of variables, so use of multivariate methods are appropriate in making use of the information from the full spectral data. PCA-based identification methods, as used with the NanoRam 785, are developed with a minimum of 20 Raman spectra, and thus provide robust models that can include variation of materials from different lots, batches and manufacturers. In this way, PCA-based methods can provide finer discrimination of materials than a correlation approach using HQI matching. This methodology can discriminate similar materials such as different types of cellulose, ethylene glycol vs. diethylene glycol, magnesium stearate vs. stearic acid, and potassium carbonate vs. potassium carbonate sesquihydrate.

The PCA-based approach, though more robust than library-matching methods, requires collection of 20 or more spectra. Method development with material samples from different batches and lots results in a representative model that spans the variability measured in the material. PCA is a data reduction technique and describes the structured spectral variance, and a PCA model is weighted towards the regions with high variance across the method spectra. Therefore, a sample with low levels of contaminants may still pass the PCA identification method if the contaminant spectrum has peaks in other regions.

Targeted multivariate identification techniques have been developed that use a "barcode" approach so that analysis is done in information-rich regions of the Raman spectrum based on the major peaks in the

spectrum.[7,8] The barcode method developed by Patel et al. used spectral regions selected from second-derivative Raman spectra as inputs into PCA analysis and effectively discriminated different strains of bacteria that have some overlap of classes when a broader spectral region was used in the PCA analysis. The advantage of the barcode approach was attributed to the removal of non-essential variances.

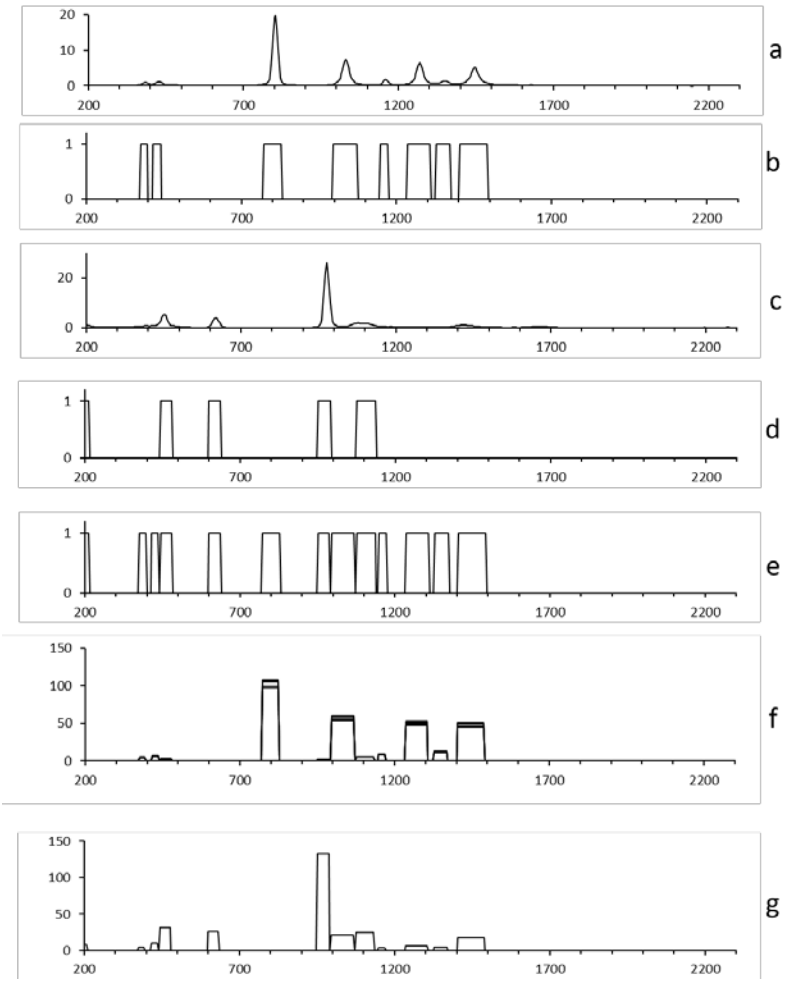
Reduced Variable Multivariate Analysis

The new multivariate identification algorithm developed for the NanoRam-1064 incorporates a variable reduction step to determine the p-value (patent pending). The p-value is used as criterion for sample identification, with values from 0 to 1, with lower p-values indicating less likelihood that the sample is the same as the selected target. The threshold for acceptance of the sample as the target material is a p-value ≥0.05.

The reduced variables of the Raman spectrum are chosen as the bands around the dominant peaks in the target material spectrum or its derivatives. In this way, nonessential features are excluded.

The variable reduction step of the RVM is illustrated in the figure below. Figure 1a represents the average of the five pre-processed spectra of the target material, cyclohexane. Fig. 1b shows eight spectral segments (regions spanned by the bars with unit height) found corresponding to the Raman peaks of the target material. Figure 1c is the spectrum of ammonium sulfate, the sample to be identified, and Figure 1d identifies five spectral segments, each representing a spectral region where a Raman peak is found for the sample, but not the target material. Figure 1e represents the sum of the segments from the target and sample, totaling 13. Having determined the m=13 spectral segments, the intensity values within each segment are summed up for each of the five target spectra and one sample spectrum, producing six spectra, each having 13 wavelength variables. The five target spectra are overlaid in Figure 1f, and the one sample spectrum in Figure 1g, where each variable is represented by a vertical bar, with its height representing the intensity and the width corresponding to the spectral segment, wherein the width of the vertical bars are only for illustration clarity and are of no computational consequence. Thus, the original 526 variables of the Raman spectrum from 176-2500 cm^{-1} are now reduced to 13 variables.

From the spectra in these reduced dimensions, the p-value is calculated using multivariate analysis. In this example, m=13, the p-value determined using the $\chi^2(13)$ distribution is 0, meaning there is an extremely low likelihood that the sample has the same composition as the target material cyclohexane, giving a FAIL result for the identification.



A systematic study using 52 compounds has shown that the Reduced Variable Multivariate (RVM) method has a higher specificity and selectivity and is more robust than the PCA-based p-value algorithm, and can be more rapidly developed using five representative spectra of the target material for each method.

The identification methods developed on the NanoRam-1064 are cross validated with all other system methods, verifying the selectivity of each method before it is put into use.

NanoRam-1064 Materials Identification Example

The NanoRam-1064 has reduced fluorescence contribution, meaning that colored samples such as Opadry (powders used in tablet coating) can be distinguished. Different types of cellulose (such as cellulose, methyl cellulose, and hydroxypropyl cellulose) have been successfully discriminated with the NanoRam 785[3] with the 20 spectra PCA-based methods, and also with the RVM identification algorithm of the NanoRam-1064. Chemically similar materials such as polysorbate 20 and polysorbate 80, even when measured through amber glass bottles, can be unambiguously identified and discriminated from each other with the NanoRam-1064 and RVM algorithm. The following table illustrates the capability of the NanoRam-1064 in identifying samples that can be a challenge for the NanoRam 785.

Sample \ Method	Opadry Clear	Opadry Green	Opadry Orange	Cellulose	Gelatin	Baby Formula	Mg Stearate	Talc ST	Corn Oil	Diesel	Polysorbate 20	Polysorbate 80	Methyl Cellulose
Opadry Clear	1	0	0	0	0	0	0	0	0	0	0	0	0
Opadry Green	0	0.999999	0	0	0	0	0	0	0	0	0	0	0
Opadry Orange	0	0	0.999998	0	0	0	0	0	0	0	0	0	0
Cellulose	0	0	0	1	0	0	0	0	0	0	0	0	0
Gelatin	0	0	0	0	0.999987	0	0	0	0	0	0	0	0
Baby Formula	0	0	0	0	0	1	0	0	0	0	0	0	0
Mg Stearate	4.14E-08	0	0	0	0	0	0.999935	0	0	0	0	0	0
Talc ST	0	0	0	0	0	0	0	0.974689	0	0	0	0	0
Corn Oil	0	0	0	0	0	0	0	0	1	0	0	0	0
Diesel	0	0	0	0	0	0	0	0	0	1	0	0	0
Polysorbate 20	0	0	0	0	0	0	0	0	0	0	1	0	0
Polysorbate 80	1.16E-12	0	0	0	0	0	0	0	0	0	0	1	0
Methyl Cellulose	1.55E-15	0	0	0	0	0	0	0	0	0	0	0	1

Conclusions

The RVM method requires only a few spectra to make a model and can be quickly developed on the NanoRam-1064. Multivariate analysis of the Raman spectra on handheld Raman instruments provides more robust methodologies for identifying samples. Both the PCA-based and RVM methods are variable reduction methods, with RVM effectively removing most of the nonessential interferences in Raman spectral models and providing even higher specificity and sensitivity than a method based on PCA. The NanoRam-1064 is more effective in differentiating samples that fluoresce, and the new RVM method has improved specificity to discriminate spectrally similar compounds.

References

1. S.R. Lowry, "Automated Spectral Searching in Infrared, Raman and Near-Infrared Spectroscopy". In: J.M. Chalmers, P.R. Griffiths, editors. Handbook of Vibrational Spectroscopy. Chichester, UK: John Wiley and Sons, 2002, Vol. 3, pp. 1948–1961.
2. R.L. McCreery, A. J. Horn, J. Spencer and E. Jefferson, Noninvasive identification of materials inside USP vials with Raman spectroscopy and a Raman spectral library, J. Pharma. Science, 87, 1-8, (1998).
3. D. Yang and R. J. Thomas, Am. Pharm. Rev., December 6, 2012, http://www.americanpharmaceuticalreview.com/Featured-Articles/126738-The-Benefits-of-a-High-Performance-Handheld-Raman-Spectrometer-for-the-Rapid-Identification-of-Pharmaceutical-Raw-Materials/
4. K. A. Bakeev and R. V. Chimenti, "Pros and cons of using correlation versus multivariate algorithms for material identification via handheld spectroscopy", European Pharmaceutical Review online White Paper, July 15, 2013. http://www.europeanpharmaceuticalreview.com/19813/whitepapers/material-identification-using-handheld-spectroscopy/
5. J.D. Rodriguez, B.J. Westenberger, L.F. Buhse, and J.F.Kauffman, Anal. Chem. 2011, 83, 4061.
6. J. Zhao, K. Frano, J. Zhou. "Reverse Intensity Correction for Spectral Library Search". Applied Spectroscopy, 2017, 71(8) 1876-1883.
7. L.S. Lawson and J.D. Rodriguez, J.D. Anal. Chem. 2016, 88, 4706–4713
8. S. Patel, W. R. Premasiri, D. T. Moir and L. D. Ziegler, J. Raman Spectrosc. 2008, 39,1660–1672

