

# Spectroscopic identification in the “real world”

Tony Davies<sup>a</sup> and Tom Fearn<sup>b</sup>

<sup>a</sup>Norwich Near Infrared Consultancy, 75 Intwood Road, Cringleford, Norwich NR4 6AA, UK

<sup>b</sup>Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK

Bankers' pensions, Government borrowing or *Spectroscopy Europe* columnists, the watchword for 2009 appears to be **controversial!** In the previous version of this column, Tony Davies (the younger) was being controversial about education and in this issue I am being controversial about one of the current applications of chemometrics to the use of spectroscopy in industry.

In particular I am unhappy about the way in which chemometrics is being applied for identity testing of pharmaceutical raw materials. The operational deployment of the method produces an unknown risk that an incoming, incorrectly labelled raw material would not be detected as an error. This is known as a “false positive”.

I need to make it clear that my collaborator (Tom) and I are not in complete agreement. Tom agrees that an error could occur but he thinks that the risk is probably quite small and the only viable alter-

native is to do nothing which is certainly unacceptable. My views are that:

- Everyone using the method should be aware and admit the potential risk of a false positive.
- Every batch in a delivery should be tested by a rapid method, all must be identified as the label compound and at least one container should be tested by a “compendium method” and positively identified as the label compound.
- Research should be undertaken to find methods for estimating the risk of a false positive occurring.

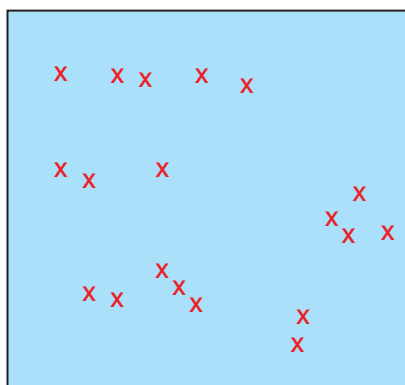
Tony Davies

## Methods for raw material identification in the pharmaceutical industry

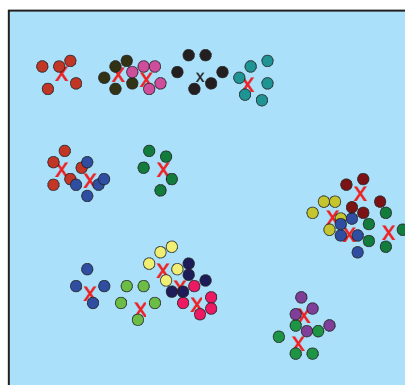
“Real world” identity testing is generally not done by the classical method of canonical variates analysis (CVA)<sup>1</sup> and

similar methods, and also not often by the modern development of CVA, soft independent modelling of class analogies (SIMCA).<sup>2</sup> SIMCA requires quite a large number of members of every group to be discriminated. The methods that are used are based on different distance measures when samples are plotted in a multi-dimensional space as described in the first article in this series.<sup>3</sup>

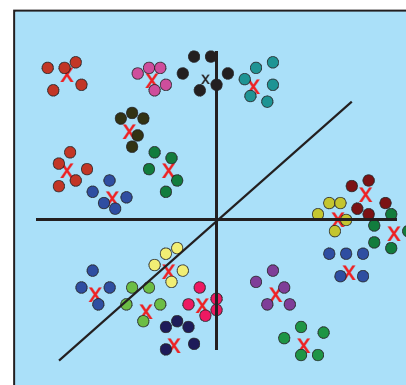
The basic steps are the same whatever method or methods are used for the classification. In Figure 1 each of the **x**s represents a chemical which is required to be present on a manufacturing site (represented by the box). The first requirement is that any chemical must be identified every time it is moved or before it is used in manufacture. This essentially means that the method must be able to distinguish between every chemical on the site, **not every known chemical in the world**. This is known as a “closed population problem”.



**Figure 1.** Each **x** represents a different chemical.

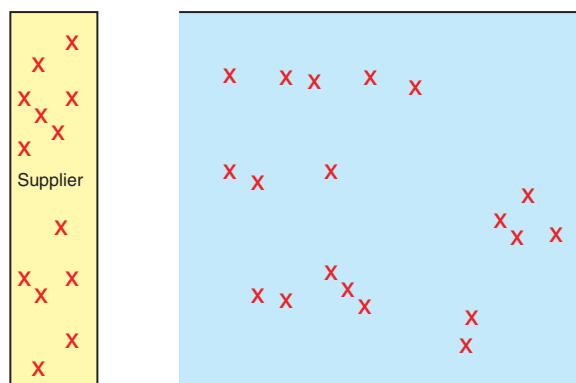


**Figure 2.** Each coloured dot represents a spectrum of a different sample of the chemical.

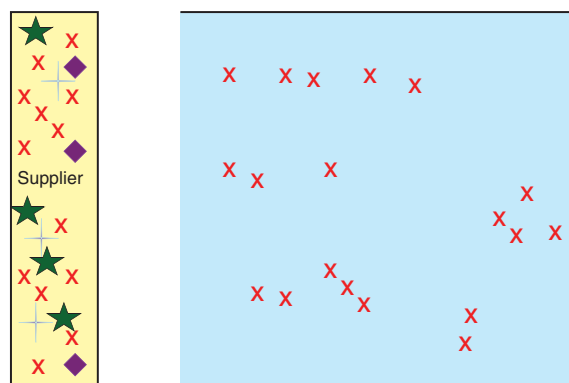


**Figure 3.** The samples now lie in a three-dimensional space that successfully separates all of them from each other.

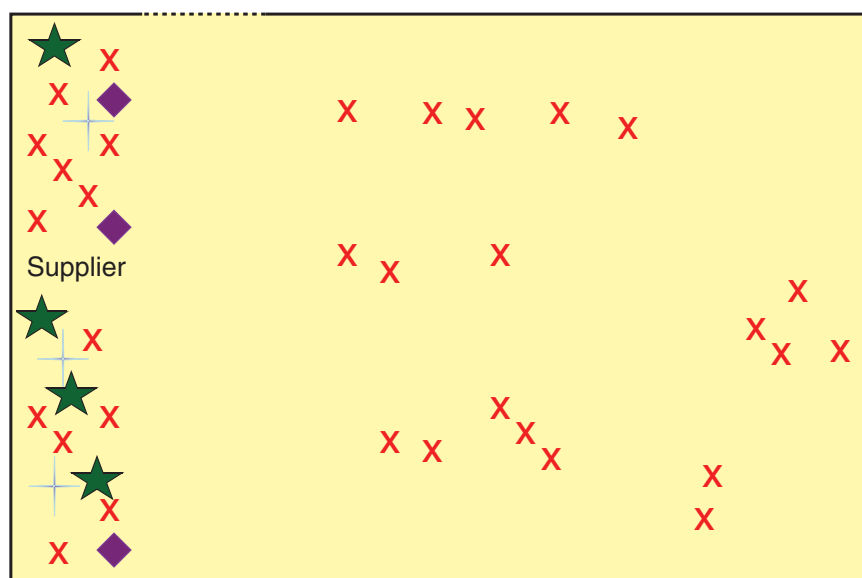
# TONY DAVIES COLUMN



**Figure 4.** A supplier and the manufacturing site. Notice that the box has been opened by the removal of one of its sides.



**Figure 5.** The supplier's site will also contain chemicals (indicated by symbols that are not **x**s) which are not required by the manufacturing site and are thus excluded from the training of the identification method.



**Figure 6.** The "closed" population now contains unknown compounds.

the different chemicals have been separated from each other. Those that appear to be over-lapping are separated in the third dimension. So now the identity of any of these chemicals can be successfully checked before they are moved or used in manufacture. **Problem solved!**

## Misuse of the method

The problem has been so successfully solved that it is often extended to the identification of incoming raw materials. The new situation is shown diagrammatically in Figure 4. The supplier will stock some of the required materials but probably not all, additional suppliers will be required. This would not be a problem if Figure 4 was an accurate representation of the supplier. A more accurate picture is shown in Figure 5; the supplier also has chemicals which are not required by the client site. However, it is common practise to use the system set up in Figure 3 to test all materials from the supplier's site. When this is done, in effect the box has been extended, Figure 6, so that the "closed" population now contains samples of unknown identity. If the supplier accidentally labels one of the un-required (and un-tested) samples as being one of the required samples then there will be an unknown risk that it will **NOT** be rejected.

It would be possible to avoid the potential problem by making Figure 6 into a proper closed sample analysis by

In order to apply chemometrics we need each of these different chemicals to be represented by spectra of a set of samples of the chemical from different batches and different suppliers. So when you see Figure 1 or similar figures you need to imagine a picture similar to Figure 2, where each sample has a number of representative spectra.

In this article we are not going to discuss, in detail, any of the methods used to tackle the problem. The methods used by individual users are often determined by their choice of spectrometer, as the software is provided by the instrument manufacturer (although some users prefer to use software from specialist software companies).

For this discussion the chemometrics will be as transparent as it is to the actual user: typically, a warehouse operative who has an NIR spectrometer on a trolley equipped with a fibre-optic probe. The operative scans a barcode on a container, then takes a spectrum of the material through the plastic bag that holds the chemical. If the software successfully correlates the sample identity with the barcode then the operator is given a green light. If it does not, a red light is illuminated and the supervisor is advised of a problem. In our case this will be represented as moving from Figure 1 to Figure 3.

Figure 3 represents a successful application of the chosen method in which all

*continued on page 31*

continued from page 27

including all the un-required samples in the identity testing training system. This could be done for one supplier but most manufacturers have many suppliers and the problem becomes unmanageable.

The European Medicines Agency (EMA) guide lines for the use of NIR spectroscopy in qualitative testing<sup>4</sup> require spectroscopists to include "*Potential challenges should be presented to the spectral reference library. These challenges should be rejected (no match). For the identification or qualification of pharmaceutical substances, relevant existing name- and structure-analogues should be included in the validation set, unless their absence is justified.*" This guidance will reduce the potential risk but, by their nature, accidents are often difficult to predict.

In the EMA guidance General requirements/Change Control and Maintenance/ (Para 4.5.4) it states: "*The use of a calibration model to analyse samples with characteristics outside of the defined scope of the method is not valid and would be considered a major GMP deficiency.*" This appears to be considered to apply only to quantitative analysis but it should be considered

as a requirement for all chemometric methods.

### What is the solution?

Most deliveries include several containers of the same chemical; some companies will take a sample from one of these containers and use a "compendium method" to establish its identity. Then they will use NIR methods to check that the material in each container is very similar to the identified material. This probably means that the risk has been transferred to human expertise and this has to be acceptable in a world that cannot be risk free.

Research is required if we are going to gain more understanding of risk analysis in relation to the problem of potential false positives in automated raw material identification.

### References

1. A.M.C. Davies and T. Fearn, *Spectrosc. Europe* **20(4)**, 18 (2008).
2. A.M.C. Davies and T. Fearn, *Spectrosc. Europe* **20(6)**, 15 (2008).
3. A.M.C. Davies and T. Fearn, *Spectrosc. Europe* **13(4)**, 22 (2001).
4. Doc. Ref. EMA/CHMP/CVMP/QWP/17760/2009 Rev 1 London, 16 February 2009.

## Complete X-Ray Spectrometer

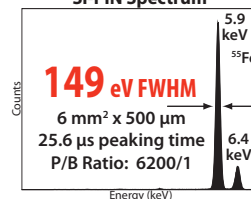


### INCLUDES

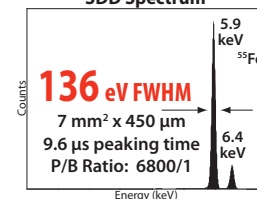
- 1 X-Ray Detector and Preamplifier
- 2 Digital Pulse Processor and MCA
- 3 Power Supply

Available with your choice of Si-PIN Detector or Silicon Drift Detector

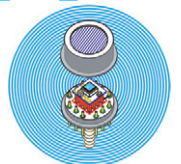
Si-PIN Spectrum



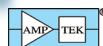
SDD Spectrum



NEW!  
Silicon Drift Detector



OEM's #1 Choice  
for XRF



Visit [www.amptek.com](http://www.amptek.com) for complete specifications  
AMPTEK Inc. sales@amptek.com www.amptek.com

**FASTLINK / CIRCLE 013 FOR FURTHER INFORMATION**