

Qualitative analysis: is it time to use the quadratic correlation approach?

Anthony M.C. Davies

*Norwich Near Infrared Consultancy, 75 Intwood Road, Cringleford, Norwich NR4 6AA, UK
td@nnirc.co.uk*

Introduction

Rose¹ introduced the use of near infrared data for quantitative analysis in 1982 using canonical variate analysis. This was followed by the publication of Mark and Tunnel² proposing the use of Mahalanobis distance as a measure of similarity in 1985. A number of different techniques are in current use. The most simple way of comparing two sets of measurements is by the correlation coefficient, r , but this does not give satisfactory performance in practical applications because variation in particle size leads to conflicting results. In 1996 Coene, Grinter and Davies³ showed that by following the work of Norris (1965)^{4,5}, who used quadratic regression (QR) to overcome particle size effects, similar success could be achieved with correlation coefficients calculated from QR results. This technique is very simple to apply and is very easily understood; it is likely that with some attention to wavelength selection it could be used for all but the most difficult identifications. Although there has been much interest in the area (Downey reviewed some of it related to food authentication⁶) there have been very few comparative studies. This is especially true of the methods used in commercial software; every provider has found it necessary to have their own unique method with very little published justification.

Perhaps now would be a good time for some comparative studies to discover if there is a need for anything more complex than QR?

Experimental

The spectra used in this explanation were measured on a Foss XDS spectrometer but the idea is applicable to spectra measured on any spectrometer. The data were originally collected at 0.5 nm intervals over the range 400 – 2498 nm but this was reduced to 1100 – 2498 nm at 2.0 nm intervals. The required calculations were conveniently carried out in an Excel (TM) spreadsheet.

The method requires that an r^2 value is calculated when the known spectrum is plotted against the unknown spectra. Normally this would be comparing the data to a linear trend line but it can work in just the same way with a quadratic trend line.

Because r^2 tends to a value of 1.0, similar samples have rather similar values. In order to make the differences between samples more apparent, a new value, named the “Similarity Index” (SI), is calculated:

$$SI = 1/(1-r^2) \quad (1)$$

The SI value can range from 1.0 to infinity.

Samples

The red dyes (Acid red 1 and Direct red 75) used for this demonstration were produced by Aldrich. The measurements were made in cells made from 35 mm film containers (see ^{ P2-21}). In

order to make large spectral differences, samples of the dyes were diluted with sulphur or “low sodium” salt (a mixture of NaCl and KCl).

Results

Spectra of the red dyes are shown in Figure 1. The results of the spreadsheet calculations are shown in Table 1 and Figures 2 and 3.

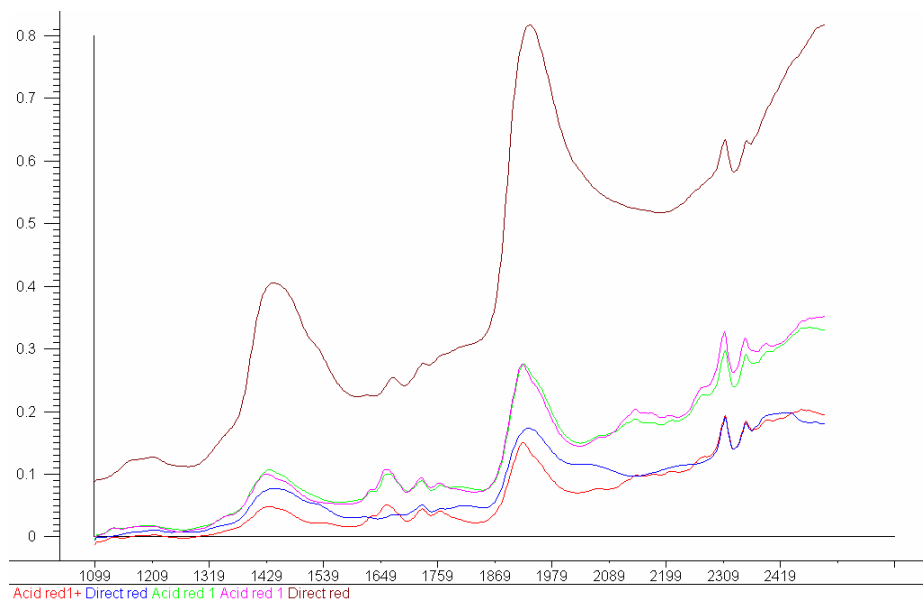


Figure 1. Spectra of red dyes before and after dilution. These spectra are identified as: Acid red (15% diluted in sulphur), Direct red (15% diluted in Sulphur), Acid red (15% diluted in NaCl/KCl), Acid red (100%) and Direct red (100%)

Table 1. Similarity Scores for different combinations of test spectra

Spectra	Linear r^2	Quadratic r^2	Linear Similarity	Quadratic Similarity
Ar1 / Ar2	0.9948	0.9951	96	102
Ar2 / Ar3	0.9912	0.9843	32	57
Dr1 / Dr2	0.9511	0.9701	10	17
Ar1 / Dr1	0.9148	0.9181	6.1	6.4

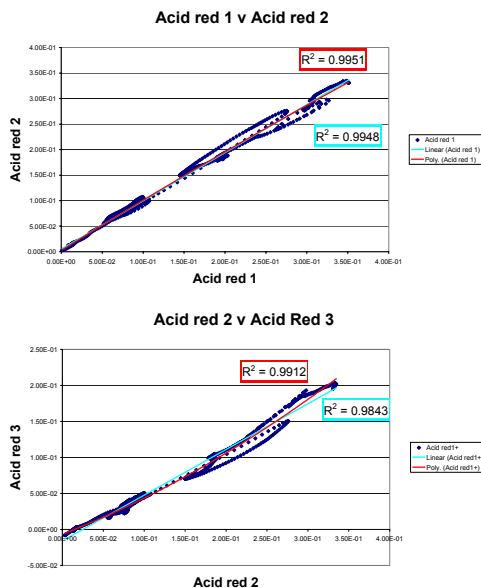


Figure 2. Comparison of very similar spectra.

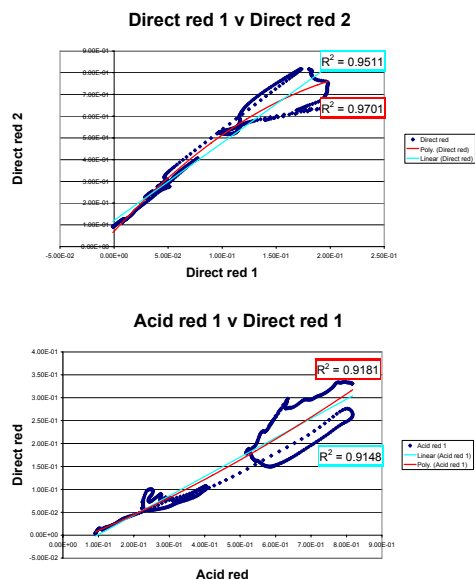


Figure 3. Comparison of less similar spectra.

Discussion

At the recent Büchi meeting in Basle (September 2002), two speakers highlighted the difficulties of utilising the frequently used, but more complex approaches to identity testing.

One speaker had a responsibility to check imported raw materials. These were often materials that he had not seen previously, there were many containers and the requirement was for a 100% inspection. His solution was to identify the contents of one container by classical IR and then use NIR to demonstrate that all other containers were similar. I think that with one authentic sample (obtained from a reliable source) he could have used QR/NIR for the complete task.

The other speaker was employed by one of the large chemical suppliers and he had the task of introducing NIR methods into the quality control system for thousands of different chemicals. The construction of libraries of identified samples for so many different chemicals is obviously an impossible task and he was trying to find alternative methods. I suggest that QR would be the answer.

The results shown in Table 1 indicate that when a sample has the same identity as the known sample then the SI value increases from the linear regression to the QR result; but when the sample is different the SI value does not change very much. This is similar to our experience in the previous work³.

Conclusions

QR is a simple technique that makes comparison of regression values useful for identity checking.

It could be used for those cases that are not amenable to the “Product library” approach.

It may be used with similar efficiency for those cases where the “Product library” approach is being used.

Setting up a QR methodology would be much less demanding than that required for the “Product library” approach.

A large scale test is required to determine the merits of the QR approach.

References

1. J. Rose, Paper 182 Pittsburg conference 1982.
2. H.Mark and D.A. Tunnell, *Anal. Chem.*, 57, 1449 (1985).
3. M.P.D. Coene, R. Grinter and A.M.C. Davies, *J. Near Infrared Spectrosc.* 4, 153–161 (1996).
4. Norris KH, Hart JR, *Proceedings of the 1963 International Symposium on Humidity and Moisture*, Volume 4, 19-25 (1965).
5. Norris KH, Hart JR, *J. Near Infrared Spectrosc.* 4, 23-29 (1996).
6. G. Downey, *J. Near Infrared Spectrosc.* 4, 47-61 (1996).