The idea behind comparison analysis using restructured near infrared and constituent data (CARNAC)

Anthony M.C. Davies
Norwich Near Infrared Consultancy, 75 Intwood Road, Cringleford, Norwich, NR4 6AA, UK

Introduction

The idea for "Comparison Analysis using Restructured Near infrared And Constituent data (CARNAC)" was formulated in 1983. However it was only through the active participation of Professor Fred McClure that it could be demonstrated in 1986 at the IDRC at Chambersburg, USA and the FT Conference in Vienna.[1]

The idea of CARNAC is that prediction of quantitative information can be derived from databases containing near infrared (NIR) and analytical data rather than through some form of regression analysis derived from that database. The realisation of CARNAC involves the combination of several ideas: database compression, database modification and similarity analysis. Through these steps it is assumed that a few samples, which are very similar to an unknown sample, can be found in a modified database. Because the database has been modified to emphasise the spectral features of the analyte it is assumed that the analyte value for the unknown sample can be estimated from the analytical values of the selected samples.

Methods

Data compression

The original work on CARNAC[1] utilises compression by Fourier transformation.[2,3] In the 1980s it was assumed that data compression would be essential to obtain an analytical result in real time but because of the vast improvements in the speed of personal computers there is no absolute need for data compression in 2001. The use of principal components (PCs) for the data compression step was demonstrated at IDRC-2000.

Database modification

Unless the database is modified by some method that emphasises the analyte, selections of samples would be identical for the same unknown sample for different analytes and all selections would be dominated by the major analyte. The method used for most of our work used the coefficients generated by running an stepwise multiple linear regression (SMLR) program for the analyte of interest. The database and the unknown sample were multiplied by the regression coefficients as a means of emphasising the analyte contribution to the spectrum. Other methods for modification of the database that have been tested on CARNAC include multiplication of the database by the spectrum of the analyte and the selection of principal components that were correlated to the analyte.

Near Infrared Spectroscopy: Proceedings of the 10th International Conference

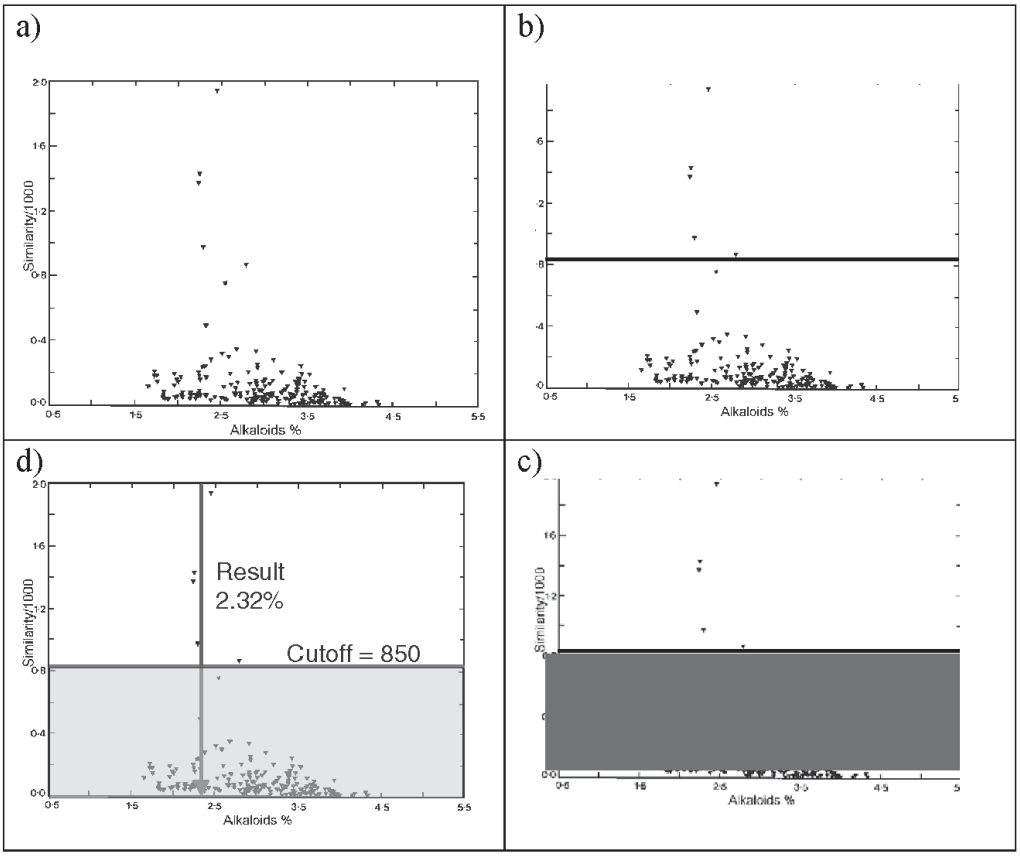

Figure 1. The main CARNAC process. (a) The SI values can be plotted against the value of the analyte for that member of the database. (b) Most samples are not similar to the unknown and they can be eliminated by drawing a horizontal line and (c) ignoring samples below it. (d) The result is calculated from a weighted average of the analyte values of the remaining samples.

Prediction of unknown samples

The modified unknown sample was compared to each member of the modified database by calculating a "similarity index"[1] (SI). The SI is calculated from the correlation coefficient (r) between the unknown sample and a member of the database The SI is calculated as:

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

The few samples with very high similarity were then selected by using a minimum cut-off value and eliminating samples identified as outliers. The predicted analyte value for the unknown sample was calculated as the weighted average of the analyte values of the selected samples. This is shown diagrammatically in Figure 1.

The logic steps are shown in Figure 2 and 3.

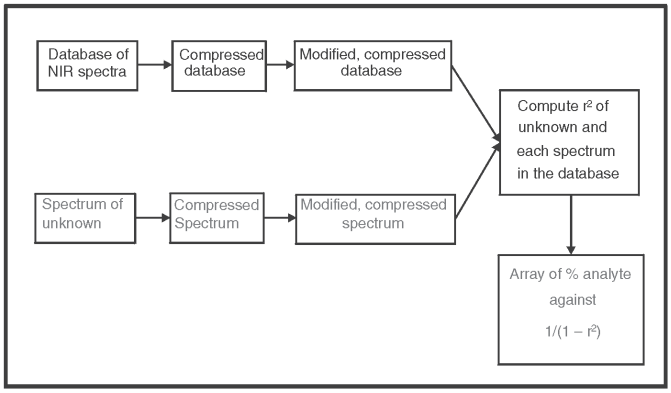

Figure 2. Logical steps for the first stage of CARNAC.

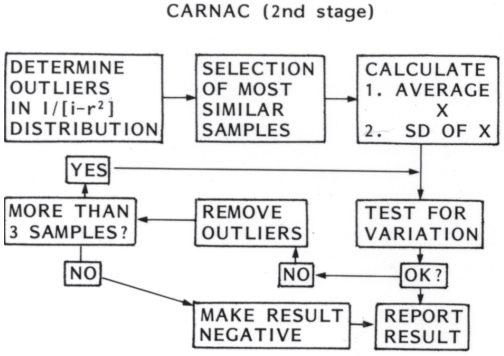

Figure 3. Logical steps for the second stage of CARNAC.

Results

Some results obtain by the CARNAC procedure are given in Figures 4 and 5.

Discussion

The most important of these results are those for caffeine in coffee. Caffeine is a minor component (1–2.5%); the method depends on the modification step to achieve these results.

When CARNAC was developed 15 years ago it was not really a practical proposition because of variation between instruments and the lack of interest in data compression by FT. My interest in CARNAC was rejuvenated by the success of a different, but related, technique developed by Shenk and Westerhaus called "LOCAL". In LOCAL a

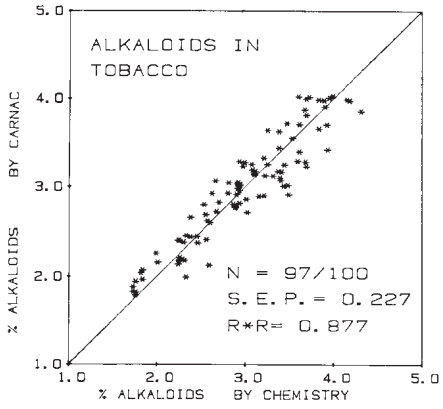

Figure 4. CARNAC results for alkaloids in tobacco.

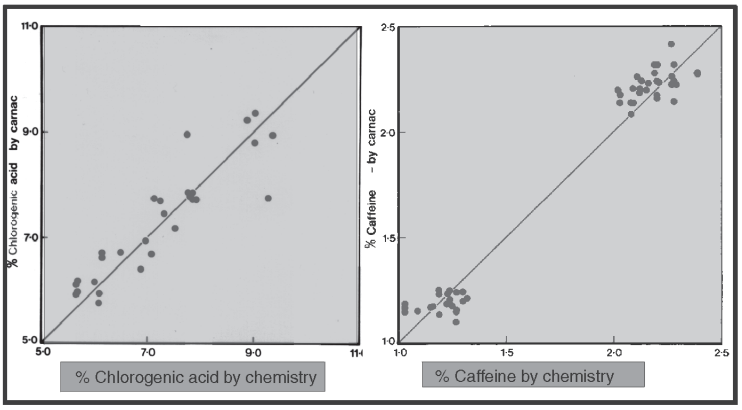

Figure 5. CARNAC results for chlorogenic acid and caffeine in coffee.

subset of samples which are similar to the spectrum of the unknown sample is selected from a compressed database and a PLS model is then constructed for these samples and used to predict the unknown sample. Thus, a new PLS calibration is determined for each unknown sample. In application, LOCAL has been successfully applied using very large databases (< 6000 in one example) and with several hundred samples being selected to form the PLS calibration set.[4–7]

Tom Fearn is collaborating with me to produce a modernisation of the CARNAC idea. A recent issue of *NIR news* contains a "Chemometric Space" article[8] by Tom, which discusses the background of these methods.

Acknowledgement

I would like to emphasise that the development of CARNAC would not have been possible without the enthusiastic collaboration of Professor Fred McClure.

References

1. A.M.C. Davies, H.V. Britcher, J.G. Franklin, S.M. Ring, A. Grant and W.F. McClure, *Mikrochim. Acta (Wien)* **I,** 61 (1988).
2. W.F. McClure, A. Hamid, F.G. Giesbrecht and W.W. Weeks, *Appl. Spectrosc.* **38,** 322 (1984).
3. W.F. McClure and A.M.C. Davies in *Analytical Applications of Spectroscopy*, Ed by C.S. Creaser and A.M.C. Davies. Royal Society of Chemistry, London, UK, p. 414 (1988).
4. J.S. Shenk and M.O. Westerhaus, *Crop Sci.* **31,** 469 (1991).
5. J.S. Shenk, P. Berzaghi and M.O. Westerhaus, *J. Near Infrared Spectrosc.* **5,** 223 (1997).
6. F.E. Barton, II, J.S. Shenk, M.O. Westerhaus and D.B. Funk, *J. Near Infrared Spectrosc.* **8,** 201 (2000).
7. P. Dardenne, G. Sinnaeve and V. Baeten, *J. Near Infrared Spectrosc*. **8,** 229 (2000).
8. T. Fearn, *NIR news* **12(3),** 10 (2001).