Introduction
In our previous column\(^1\) we introduced some distance statistics that have been used for comparing spectra. These calculations provide univariate answers from multivariate data in a single step. This may be adequate for some problems but often we need to employ some multivariate mathematics before the reduction to a univariate answer.

This column is an introduction to the first method, which was invented long before chemometrics by R.A. Fisher; some seventy years ago! Canonical Variates Analysis (CVA)\(^2\) has been one of my favourite examples of chemometrics because it often requires the use of a compression technique (PCA or FFT for example) before it can be applied and I think it helps students to understand the need to know the essential properties of the different tools in the chemometric toolbox.

Tony Davies

Groups
In multivariate analysis of spectroscopic data it is very unusual to compare an unknown with a single spectrum of a known sample. It is normal to collect spectra from several examples of the same sample into a group and compare the unknown spectrum with the group. This is because when we make measurements there will always be some variation between different examples and we need to have information about the variability of the group. In fact instrument noise ensures that spectra of the same example measured on the same instrument will have some variability.

Canonical variate analysis
The CVA technique has similarities with PCA in that the multivariate data is submitted to the program which computes new variables and values (scores) for each sample and each of the new variables. In PCA the new variables are principal components, while in CVA they are canonical variates. Where they differ is in how the new variables are selected. PCA is not given any information about groups and group membership, it is just required to compute new variables to maximise the variability of the scores for the whole data set. CVA is given information about groups and group membership and the requirement is that new variables will minimise the within-group variation while maximising the between-group variation. As shown in Figure 1(a) and (b), the within-group
variance, $W_g$, is a pooled result from all the groups being considered. When the groups have different variability, as in Figure 1, $W_g$ is a compromise, but the approach often works well in spite of this.

$B_g$ is the between group variance and the criterion to be maximised is the ratio $B_g/W_g$.

To apply CVA, the number of input variables must be less than (in reality considerably less than) the number of samples. This presents a difficulty with spectroscopic data that usually has a large number of variables (wavelengths or wavenumbers). The possible solutions are to discard the majority of datapoints or (as mentioned earlier) to use some form of compression to retain most of the information in the original data but compressed into fewer variables. The most obvious of these is PCA.

One difference between PCA and CVA is that the transformation from original variables to scores in PCA is a simple rotation in which the axes remain mutually orthogonal. In CVA the angles between the axes and the scaling of the axes changes so that the elliptical shape describing within-group variability (and corresponding to $W_g$) becomes a circular or spherical shape. In consequence, measuring using Euclidean distances in CV-space corresponds to using Mahalanobis distance in the original spectral space, and classification using the CVA approach is equivalent to classification using Mahalanobis distance (see our frequently referenced book\(^3\) for a description of Mahalanobis distance).

**Examples**

a) Two groups

With two groups we need to find only one CV. The example was mentioned in the first "Chemometric Column" in *Spectroscopy World*\(^4\) and again in an early "Tony Davies Column"\(^5\). It involves the separation of regular and decaffeinated instant coffee samples from their NIR spectra. The spectra, Figure 2(a), show no separation. The spectra contained 700 datapoints and at that time our best PCA program would accept only 50 variables. One way to utilise this program was to reduce the number of wavelength variables by averaging successive 14 data points leaving 50 variables so that these could be used as the input data. Then the first ten PCs were used as the input data to a CVA program. The result obtained from the CVA using the first ten PCs is shown Figure 2(c). Nowadays, PCA programs accept much larger numbers
of input variables but at the time it was a
demonstration of the value of compres-
sion by FT and also the value of using
chemometrics tools for their designed
application.

b) Several groups
This example is from some later
work to attempt to identify the botani-
cal origin of honey samples from their
NIR spectra. Second derivative spec-
tra were used as the input data to a
PCA program and the first ten or first
fifteen PCs were used as the input
data to a CVA program. This work was
very much a preliminary study because
there were very few samples available.
The work was validated using cross-
validation leaving out one sample at a
time and recalculating the PCs as well
as the CVs. Although the best separ-
tions used three CVs it is easier to look
at just the first two. Figure 3(a) shows
the PCA result for one particular sample
removed and Figure 3(b) shows the
CVA result obtained from this data.
These graphs are an excellent example
for demonstrating the superior power
of CVA compared to PCA for separat-
ing similar samples. Many people use
PCA for identification; it does work in
many cases but it only works by acci-
dent because the variability in the data
is related to the differences between
samples. PCA is an “unsupervised
method” it cannot make use of the
information about group membership
to improve the separation.

Coming soon
In our next column we will discuss the
more recently developed method of
SIMCA and discuss the other factors
which must be taken into considera-
tion before discrimination decisions can be
made.

References
1. A.M.C. Davies and T. Fearn,
Spectroscopy Europe 20(2), 15
(2008).
(1936).
3. T. Næs, T. Isaksson, T. Fearn and
T. Davies, A User-Friendly Guide
to Multivariate Calibration and
Classification. NIR Publications,
Chichester (2002).
4. A.M.C. Davies, Spectroscopy World
5. A.M.C. Davies, Spectroscopy Europe
5(6), 30 (1993).
6. A.M.C. Davies, B. Radovic, T. Fearn
and E. Anklam, J. Near Infrared
Spectrosc. 10, 121–135 (2002).