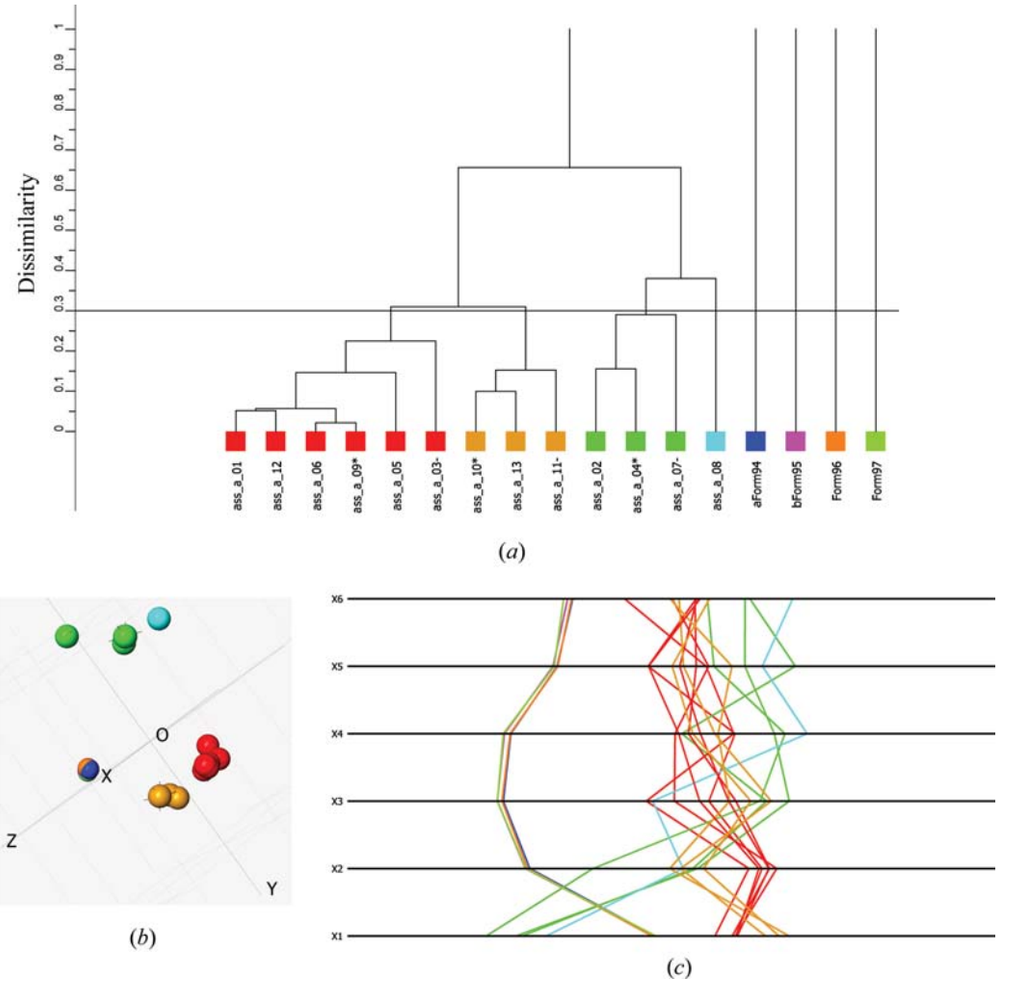


Figure 3.8.10
The aspirin data including data from five amorphous samples. (*a*) The resulting dendrogram and (*b*) the corresponding MMDS plot. (*c*) The parallel-coordinates plot.

$a = 7.14$, $b = 7.65$, $c = 5.83$ Å with $Z = 4$; between 357 and 398 K it crystallizes in the tetragonal space group $P\bar{4}2_1m$ with $a = 5.719$, $c = 4.932$ Å, $Z = 2$, and above 398 K it transforms to the cubic space group $Pm\bar{3}m$ with $a = 4.40$ Å and $Z = 1$. PXRD data containing 75 powder patterns taken at intervals of 3 K starting at 203 K using a D5000 Siemens diffractometer and Cu $K\alpha$ radiation with a 2θ range of 10–100° were used (Herrmann & Engel, 1997). Fig. 3.8.11(*a*) shows the data in the 2θ range 17–45°.

The visualization of these data following cluster analysis is shown in Fig. 3.8.11(*b*) using an MMDS plot on which has been superimposed a line showing the route followed by the temperature increments. The purple line follows the transition from a mixture of forms IV and V at low temperature (red) through form IV (yellow), form II (blue) and finally form I at high temperature (green). This is an elegant and concise representation of the data in a single diagram.

3.8.7. Quantitative analysis with high-throughput PXRD data without Rietveld refinement

Since mixtures are so common in high-throughput experiments, and indeed in many situations with multiple data sets, it is useful to have a method of automatic quantitative analysis. The quality of data that results from high-throughput crystallography makes it unlikely that an accuracy better than 5–10% can be achieved but, nonetheless, the identification of mixtures can be carried out by whole-profile matching. First a database of N pure phases is created, or, if that is not possible, then the most representative patterns with appropriate safeguards can be used. Assume that there is a sample pattern, S, which is considered to be a mixture of up to N components. S comprises m data points, $S_1, S_2, \ldots, S_m$. The N patterns can be considered to make up fractions $p_1, p_2, p_3, \ldots, p_N$ of the sample pattern. The best possible combination of the database patterns to fit the sample pattern is required. A system of linear equations can be constructed in which x_{11} is measurement point 1 of pattern 1 *etc.*:

$$\begin{aligned} x_{11}p_1 + x_{12}p_2 + x_{13}p_3 + \ldots + x_{1N}p_N &= S_1, \\ x_{21}p_1 + x_{22}p_2 + x_{23}p_3 + \ldots + x_{2N}p_N &= S_2, \\ &\vdots \\ x_{m1}p_1 + x_{m2}p_2 + x_{m3}p_3 + \ldots + x_{mN}p_N &= S_m. \end{aligned} \tag{3.8.26}$$

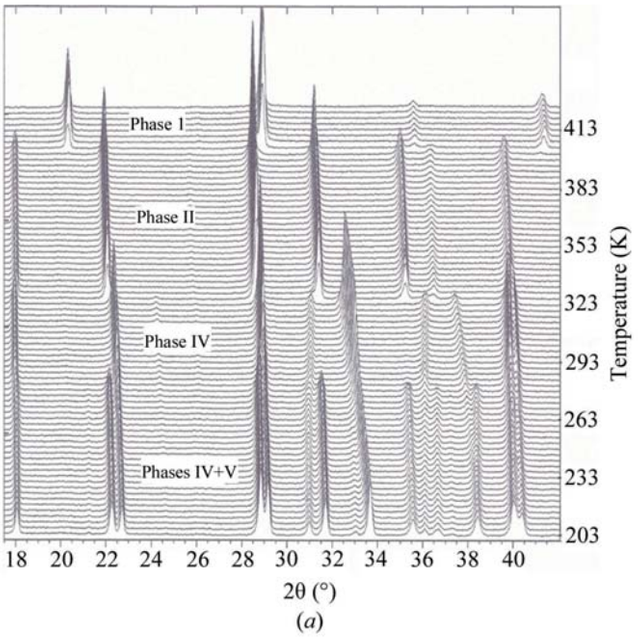
Writing these in matrix form, we get

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1N} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mN} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{bmatrix} \tag{3.8.27}$$

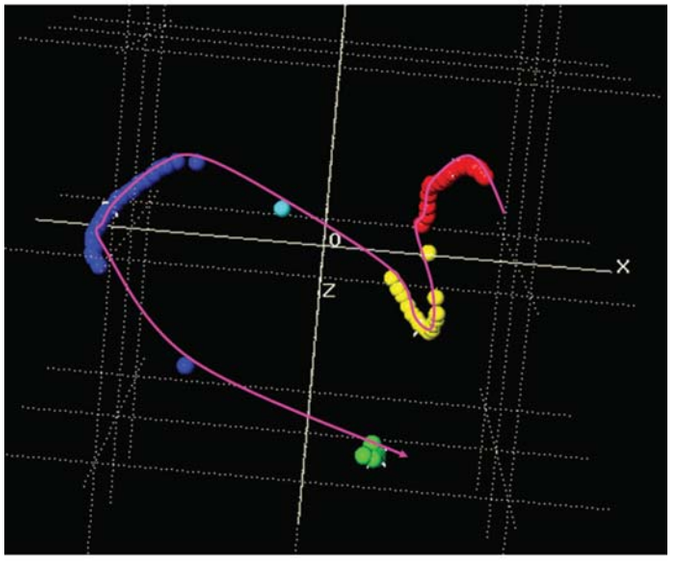
or

$$\mathbf{xp} = \mathbf{S}. \tag{3.8.28}$$

A solution for S that minimizes



(*a*)



(*b*)

Figure 3.8.11
Ammonium nitrate phase transitions. (*a*) The raw powder data measured between 203 and 425 K. Reproduced with permission from Herrmann & Engel (1997). Copyright (1997) John Wiley and Sons. (*b*) The MMDS plot. The purple line follows the temperature change from 203 to 425 K.

$$\chi^2 = \left|\mathbf{xp} - \mathbf{S}\right|^2. \tag{3.8.29}$$

is required. Since $N \ll m$, the system is heavily overdetermined, and least-squares or singular value decomposition can be used to solve (3.8.29) for the fractional percentages arising from the scattering power of the component mixtures, $s_1, s_2, \ldots, s_N$. The values of s can be used to calculate a weight fraction for that particular phase provided that the atomic absorption coefficients are known, and this in turn requires the unit-cell dimensions and cell contents, but not the atomic coordinates (Smith *et al.*, 1988; Cressey & Schofield, 1996). The general formula for the weight fraction of component n in a mixture comprising N components is (Leroux *et al.*, 1953)

$$c_n = p_n \frac{\mu^*}{\mu_n^*}, \tag{3.8.30}$$

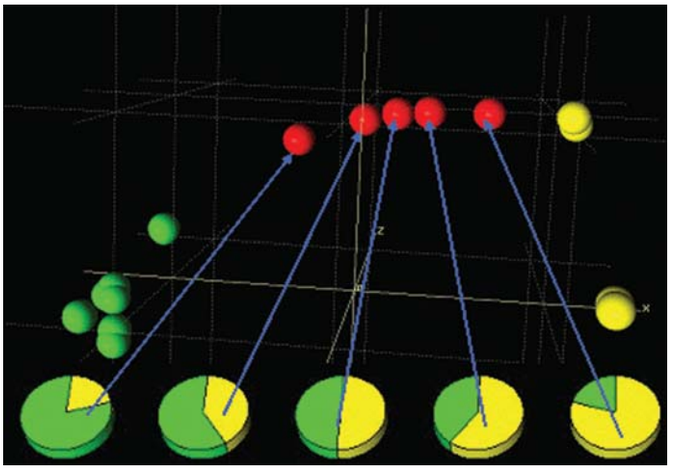


Figure 3.8.12
Identifying mixtures using lanthanum strontium copper oxide and caesium thiocyanate diffraction data taken from the ICDD Clay Minerals database. The green spheres represent pure phases of lanthanum strontium copper oxide and the yellow pure caesium thiocyanate. The red spheres represent mixtures of the two in the relative proportions of lanthanum strontium copper oxide/caesium thiocyanate 80/20, 60/40, 50/50, 40/60 and 20/80 in an arc commencing on the left-hand side of the diagram. The pie charts give the results of an independent quantitative calculation in which lanthanum strontium copper oxide and caesium thiocyanate have been included as pure phases in a reference database.

where

$$\mu^* = \sum_{j=1}^{N} c_j \mu_j^* \tag{3.8.31}$$

and

$$\mu_j^* = \mu_j / \rho_j, \tag{3.8.32}$$

where μ_j is the atomic X-ray absorption coefficient and ρ_j is the density of component j. For polymorphs, the absorption coefficients are sufficiently close and the method sufficiently approximate that the effects of absorption can be ignored.

3.8.7.1. *Example: inorganic mixtures*

As an example, a set of 19 patterns from set 78 of the ICDD database for inorganic compounds (ICDD, 2018) was imported into *DIFFRAC.EVA*. To this was added some simulated mixture data generated by adding the patterns for lanthanum strontium copper oxide and caesium thiocyanate in the proportions 80/20, 60/40, 50/50, 40/60 and 20/80. Two calculations were performed: an analysis without the pure-phase database and a second where the pure phases of lanthanum strontium copper oxide and caesium thiocyanate were present.

The results are shown in Fig. 3.8.12. In the MMDS plot the green spheres represent pure lanthanum strontium copper oxide while the yellow are pure caesium thiocyanate. The red spheres represent mixtures of the two. The latter form an arc between the green and yellow clusters. The distance of the spheres representing mixtures from the lanthanum strontium copper oxide and caesium thiocyanate spheres gives a semi-quantitative representation of the mixture contents. Running the analysis in quantitative mode gives the pie charts also shown in Fig. 3.8.12; they reproduce exactly the relative proportions of the three components.

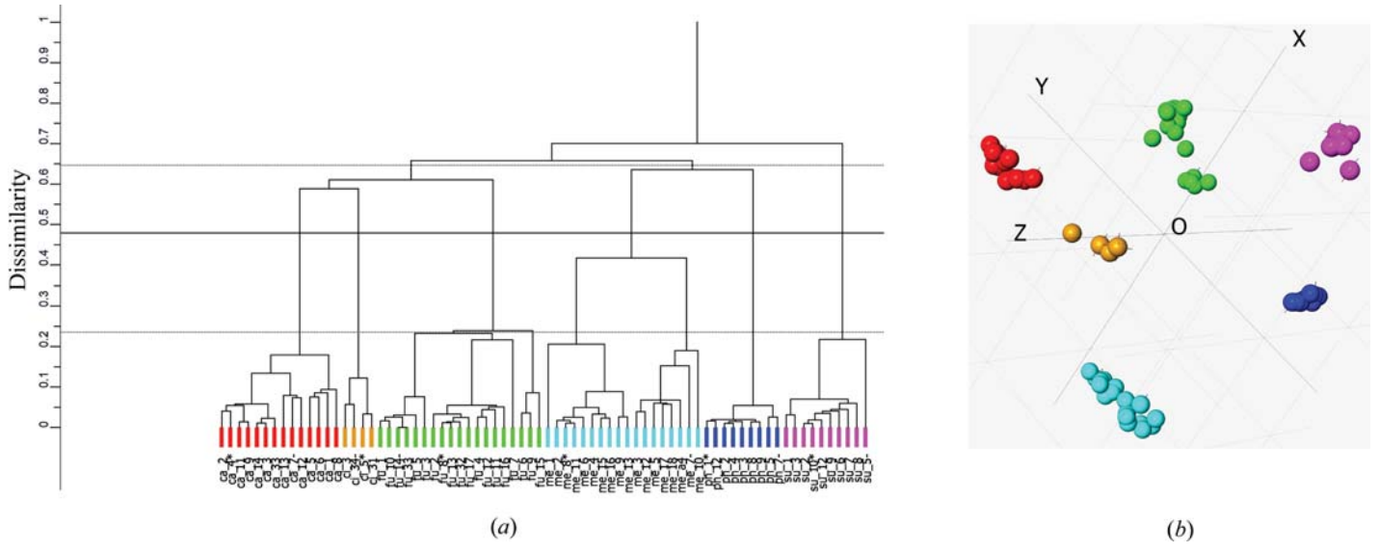


Figure 3.8.13
(*a*) The dendrogram generated from 74 Raman spectra without background corrections applied. Labelling from the left-hand side, the red samples are carbamazepine, the orange are cimetidene, the green are two forms of furosemide, the light blue is mefenamic acid, the dark blue is phenilbutazone and the purple at the right-hand side is sulfamerazine. (*b*) The MMDS plot. The sphere colours are taken from the dendrogram. This representation shows clearly discrete clusters in correspondence with those generated by the dendrogram.

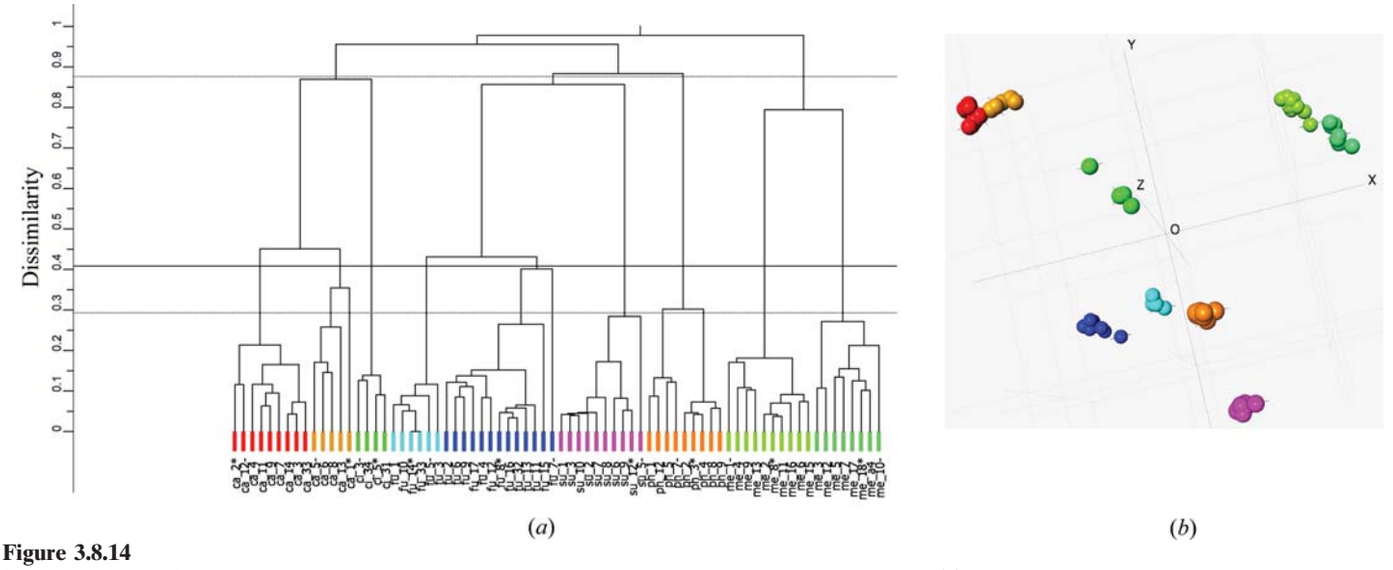


Figure 3.8.14
Clustering the 74 Raman spectra without background corrections applied using first-derivative data. (*a*) The dendrogram. Labelling from the left-hand side, the red and orange entries are carbamazepine; the green are cimetidene; the light blue and dark blue are two forms of furosemide; the purple are sulfamerazine; the brown are phenilbutazone and the right-hand light and dark green are two forms of mefenamic acid. (*b*) The MMDS plot. The clusters are well defined but the orange and red (both carbamazepine) are very close to each other.

For further details of this method with organic samples, see Dong *et al.* (2008).

3.8.8. Using spectroscopic data

There is no reason why the methodology described in this chapter cannot be used for other 1D data sets, *e.g.* Raman, IR, NMR and near-IR spectroscopies, although different data pre-processing is usually required. Raman spectroscopy is well suited to high-throughput screening: good-quality spectra can be collected in a few minutes, and sample preparation is straightforward and flexible, although the resulting spectra are not always as distinct as the PXRD equivalents (Mehrens *et al.*, 2005; Boccaleri *et al.*, 2007).

As an example we show the results of cluster analysis carried out on samples of carbamazepine, cimetidene, furosemide, mefenamic acid, phenilbutazone and sulfamerazine using Raman spectroscopy. A total of 74 samples were measured on a LabRam HR-800/HTS-Multiwell spectrometer at room temperature, equipped with a backscattering light path system of a light-emitting diode laser (785 nm, 300 mW) as an excitation source and an air-cooled charge-coupled device detector. A 20-fold superlong working distance objective lens was used to collect the backscattered light. The spectra were acquired with 5.84 cm^{-1} spectral width and at least 30 s exposure (Kojima *et al.*, 2006). The spectra had backgrounds subtracted but no other corrections were carried out.

The initial clustering is shown in Fig. 3.8.13(*a*) with the default cut level in the dendrogram. There are six clusters: labelling from the left-hand side, the red are three polymorphs of carbamazepine; the orange are cimetidene; the green cluster contains three polymorphs of furosemide; the light blue contains three poly-