Cluster Analysis in Crystallography

Gordon Barr, Wei Dong, Christopher Gilmore, *Department of Chemistry, University of Glasgow, Glasgow, UK.* E-mail: gbarr@chem.gla.ac.uk

Cluster analysis is a well established tool in statistics, but one that is used surprisingly little in crystallography despite its considerable potential. We have established its use in several diverse areas of crystallography, in particular:

1. Pattern matching in powder X-ray diffraction: High throughput screening experiments designed to search for polymorphs and salts of drug candidates use PXRD to characterize the results, and this produces large quantities of data. We show how pattern matching methods based on appropriate correlation coefficients can be used in conjunction with clustering calculations to classify patterns automatically [1,2].
2. Databases: Database searching using the Cambridge Structural Database (CSD) [3] can produce thousands of 'hits' if a simple fragment is used, and as a result processing and interpreting the results becomes a considerable task. Cluster analysis using dendrograms, metric multidimensional scaling and suitable visualization tools can reduce the workload to a few hours of computer time with minimal user intervention.
3. Indexing powder patterns: In difficult indexing problems, it is possible to produce a large number of potential unit cells with figures of merit that are only marginally useful. Cluster analysis can be useful here, especially when self-organizing maps are utilised.

[1] Gilmore C.J., Barr G., Paisley J., *J. Appl. Cryst.*, 2004, **37**, 231-242. [2] Barr G., Dong W., Gilmore C.J., *J. Appl. Cryst.,* 2004, **37**, 243-252. [3] Allen F.H., Motherwell W.D.S., Acta Cryst., 2002, B**58**, 407-422.

Keywords: cluster analysis, pattern matching, databases