

3.8. DATA CLUSTERING AND VISUALIZATION

References

- Alvarez, A. J., Singh, A. & Myerson, A. S. (2009). *Polymorph screening: comparing a semi-automated approach with a high throughput method*. *Cryst. Growth Des.* **9**, 4181–4188.
- Asimov, D. (1985). *The grand tour: a tool for viewing multidimensional data*. *SIAM J. Sci. Stat. Comput.* **6**, 128–143.
- Barr, G., Cunningham, G., Dong, W., Gilmore, C. J. & Kojima, T. (2009). *High-throughput powder diffraction V: the use of Raman spectroscopy with and without X-ray powder diffraction data*. *J. Appl. Cryst.* **42**, 706–714.
- Barr, G., Dong, W. & Gilmore, C. J. (2004a). *High-throughput powder diffraction. II. Applications of clustering methods and multivariate data analysis*. *J. Appl. Cryst.* **37**, 243–252.
- Barr, G., Dong, W. & Gilmore, C. J. (2004b). *High-throughput powder diffraction. IV. Cluster validation using silhouettes and fuzzy clustering*. *J. Appl. Cryst.* **37**, 874–882.
- Barr, G., Dong, W. & Gilmore, C. J. (2004c). *PolySNAP: a computer program for analysing high-throughput powder diffraction data*. *J. Appl. Cryst.* **37**, 658–664.
- Barr, G., Dong, W. & Gilmore, C. J. (2009). *PolySNAP3: a computer program for analysing and visualizing high-throughput data from diffraction and spectroscopic sources*. *J. Appl. Cryst.* **42**, 965–974.
- Barr, G., Dong, W., Gilmore, C. & Faber, J. (2004). *High-throughput powder diffraction. III. The application of full-profile pattern matching and multivariate statistical analysis to round-robin-type data sets*. *J. Appl. Cryst.* **37**, 635–642.
- Barr, G., Gilmore, C. J. & Paisley, J. (2004). *SNAP-ID: a computer program for qualitative and quantitative powder diffraction pattern analysis using the full pattern profile*. *J. Appl. Cryst.* **37**, 665–668.
- Boccaleri, E., Carniato, F., Croce, G., Viterbo, D., van Beek, W., Emerich, H. & Milanesio, M. (2007). *In situ simultaneous Raman/high-resolution X-ray powder diffraction study of transformations occurring in materials at non-ambient conditions*. *J. Appl. Cryst.* **40**, 684–693.
- Bruker (2018). *DIFFRAC.EVA: software to evaluate X-ray diffraction data*. Version 4.3. <https://www.bruker.com/eva>.
- Butler, B. M., Sila, A., Nyambura, K. D., Gilmore, C. J., Kourkoumelis, N. & Hillier, S. (2019). *Pre-treatment of soil X-ray powder diffraction data for cluster analysis*. *Geoderma*, **337**, 413–424.
- Calinški, T. & Harabasz, J. (1974). *A dendritic method for cluster analysis*. *Commun. Stat.* **3**, 1–27.
- Carroll, J. D. & Chang, J. J. (1970). *Analysis of individual differences in multidimensional scaling via n-way generalization of 'Eckhart-Young' decomposition*. *Psychometrika*, **35**, 283–319.
- Cook, D. & Swayne, D. F. (2007). *Interactive and Dynamic Graphics for Data Analysis with R and GGobi*. New York: Springer.
- Cox, T. F. & Cox, M. A. A. (2000). *Multidimensional Scaling*, 2nd ed. Boca Raton: Chapman & Hall/CRC.
- Crawley, M. J. (2007). *The R Book*. Chichester: Wiley.
- Cressey, G. & Schofield, P. F. (1996). *Rapid whole-pattern profile-stripping method for the quantification of multiphase samples*. *Powder Diff.* **11**, 35–39.
- Dong, W., Gilmore, C., Barr, G., Dallman, C., Feeder, N. & Terry, S. (2008). *A quick method for the quantitative analysis of mixtures. I. Powder X-ray diffraction*. *J. Pharm. Sci.* **97**, 2260–2276.
- Donoho, D. L. & Johnstone, I. M. (1995). *Adapting to unknown smoothness via wavelet shrinkage*. *J. Am. Stat. Assoc.* **90**, 1200–1224.
- Everitt, B. S., Landau, S. & Leese, M. (2001). *Cluster Analysis*, 4th ed. London: Arnold.
- Gilmore, C. J., Barr, G. & Paisley, W. (2004). *High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles*. *J. Appl. Cryst.* **37**, 231–242.
- Goodman, L. A. & Kruskal, W. H. (1954). *Measures of association for cross-classifications*. *J. Am. Stat. Assoc.* **49**, 732–764.
- Gordon, A. D. (1981). *Classification*, 1st ed., pp. 46–49. London: Chapman and Hall.
- Gordon, A. D. (1999). *Classification*, 2nd ed. Boca Raton: Chapman and Hall/CRC.
- Gower, J. C. (1966). *Some distance properties of latent root and vector methods used in multivariate analysis*. *Biometrika*, **53**, 325–328.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes Problems*. Oxford University Press.
- Herrmann, M. J. & Engel, W. (1997). *Phase transitions and lattice dynamics of ammonium nitrate*. *Propellants Explos. Pyrotech.* **22**, 143–147.
- Husson, F. & Pagès, J. (2006). *INDSCAL model: geometrical interpretation and methodology*. *Comput. Stat. Data Anal.* **50**, 358–378.
- ICDD (2018). *The Powder Diffraction File*. International Centre for Diffraction Data, 12 Campus Boulevard, Newton Square, Pennsylvania 19073-3273, USA.
- Inselberg, A. (1985). *The plane with parallel coordinates*. *Vis. Comput.* **1**, 69–91.
- Inselberg, A. (2009). *Parallel Coordinates. Visual multidimensional geometry and its applications*. New York: Springer.
- Kaufman, L. & Rousseeuw, P. J. (1990). *Finding Groups in Data*. New York: Wiley.
- Kojima, T., Onoue, S., Murase, N., Katoh, F., Mano, T. & Matsuda, Y. (2006). *Crystalline form information from multiwell plate salt screening by use of Raman microscopy*. *Pharm. Res.* **23**, 806–812.
- Kruskal, J. B. (1956). *On the shortest spanning subtree of a graph and the traveling salesman problem*. *Proc. Am. Math. Soc.* **7**, 48–50.
- Lance, G. N. & Williams, W. T. (1967). *A general theory of classificatory sorting strategies. I. Hierarchical systems*. *Comput. J.* **9**, 373–380.
- Leroux, J., Lennox, D. H. & Kay, K. (1953). *Direct quantitative X-ray analysis by diffraction absorption technique*. *Anal. Chem.* **25**, 740–743.
- Mehrens, S. M., Kale, U. J. & Qu, X. (2005). *Statistical analysis of differences in the Raman spectra of polymorphs*. *J. Pharm. Sci.* **94**, 1354–1367.
- Milligan, G. W. & Cooper, M. C. (1985). *An examination of procedures for determining the number of clusters in a data set*. *Psychometrika*, **50**, 159–179.
- Nelder, J. A. & Mead, R. (1965). *A simplex method for function minimization*. *Comput. J.* **7**, 308–313.
- O'Brien, L. E., Timmins, P., Williams, A. C. & York, P. (2004). *Use of in situ FT-Raman spectroscopy to study the kinetics of the transformation of carbamazepine polymorphs*. *J. Pharm. Biomed. Anal.* **36**, 335–340.
- Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*, pp. 144–148. Boston: Birkhäuser.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes*. 3rd ed. Cambridge University Press.
- Rousseeuw, P. J. (1987). *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. *J. Comput. Appl. Math.* **20**, 53–65.
- Sato, M., Sato, Y. & Jain, L. C. (1966). *Fuzzy Clustering Models and Applications*. New York: Physica-Verlag.
- Smith, D. K., Johnson, G. G. & Wims, A. M. (1988). *Use of full diffraction spectra, both experimental and calculated, in quantitative powder diffraction analysis*. *Aust. J. Phys.* **41**, 311–321.
- Smrčok, L., Ďurík, M. & Jorík, V. (1999). *Wavelet denoising of powder diffraction patterns*. *Powder Diff.* **14**, 300–304.
- Spearman, C. (1904). *The proof and measurement of association between two things*. *Am. J. Psychol.* **15**, 72–101.
- Storey, R., Docherty, R., Higginson, P., Dallman, C., Gilmore, C., Barr, G. & Dong, W. (2004). *Automation of solid form screening procedures in the pharmaceutical industry—how to avoid the bottlenecks*. *Crystallogr. Rev.* **10**, 45–56.
- Wegman, E. J. (1990). *Hyperdimensional data analysis using parallel coordinates*. *J. Am. Stat. Assoc.* **85**, 664–675.